PATENT
ATTORNEY DOCKET NO.: 056291-5073

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re PATENT APPLICATION of:	)	
	)	
ROSAMOND et al.	)	
	)	
U.S. Application No.: 10/069,062	)	Group Art Unit: 1645
	)	
Filed: February 21, 2002	)	Examiner: Baskar, P.
	)	
FOR: PROTEIN	)	

Commissioner for Patents
U.S. Patent and Trademark Office
2011 South Clark Place
Customer Window
Crystal Plaza Two, Lobby, Room 1B03
Arlington, VA 22202

Sir:

DECLARATION UNDER 35 U.S.C. § 1.132
Of INVENTOR JOHN ROSAMOND

1. I am John Rosamond, currently Director of Informatics in Infection Discovery at AstraZeneca R&D Boston. I have a First Class Honours degree in Biochemistry and a D. Phil. in Microbiology from the University of Oxford, UK. I have over thirty years research experience in microbial molecular biology, genetics and biochemistry in both academic and pharmaceutical industry settings.

2. I have read the examiners report for this patent application (Article Unit 1645) from which I understand that the examiner questions whether a person skilled in the art would know how to introduce mutations into the ERG8 sequence while retaining biological activity, without excessive and undue experimental work.

3. A person skilled in the art would be able to use one of several algorithms to align the ERG8 protein sequences from *Candida albicans* (hereinafter C. albicans) and *Saccharomyces cerevisiae* (hereinafter S. cerevisiae) based on information available prior

to the filing dates associated with this application. One example of such an alignment is shown in Figure 1. From such an alignment, the person skilled in the art would be able to identify regions of contiguous sequence that are conserved in both proteins, as exemplified by the regions of the proteins shown in bold italicized text in Figure 1. The conservation of such regions or domains would be known by this person as likely to play a key role in the enzymatic activity of the protein, for example by making key contributions to the 3-dimensional structure of the active site. Consequently, a person skilled in the art would recognize that these regions were unlikely to be able to accommodate changes to the amino acid sequence without affecting the biological function.

4. From the same alignment, a person skilled in the art would recognize regions that showed less overall conservation as being those parts of the protein that could potentially accommodate mutation without loss of biological function. On the basis of published data comparing other functional homologs from C. albicans and S. cerevisiae, (for example Sherlock et al., (1994) Molecular & General Genetics 245, 716-723; Nolan & Rosamond, (1996) Gene 183, 159-165) it would be known to a person skilled in the art, that such regions are typically found at the N- and C-termini of the proteins. Analysis of the aligned ERG8 proteins (Figure 1) reveals that these proteins have relatively little identity beyond residue 385 of the C. albicans protein. This region would be seen to provide scope for deletion or a series of point mutations that would be likely to retain biological function.

5. Using the cognate DNA sequence for the ERG8 gene, the person skilled in the art would be able to design primers that could be used to amplify the ERG8 gene by PCR. The primers could be further designed to modify a specific amino-acid residue in the C-terminal region or to engineer the deletion of the residues downstream of amino acid 385. The amplified product would be ligated into a suitable plasmid vector, which would be cloned, then transformed into a strain of bacterium to express the product of the ERG8 gene. Function of the mutated ERG8 gene on the plasmid would be assessed by the ability to generate active phosphomevalonate kinase using any one of several well known

assays for detecting the change in ATP and ADP levels that represent the activity of PMK in the presence of phosphomevalonate. The specification of this application discusses these assays for PMK activity on pages 11 and 12 of the as-filed application. This would allow rapid identification of mutations in ERG8 that retained function, but which differed from the original wild-type sequence either by a single mutation, or by deletion of the non-conserved C-terminal region.

6. In addition to a gross deletion of a region of the ERG8 protein, as described above, a person skilled in the art would recognize that strains carrying multiple point mutations in the ERG8 gene could be identified rapidly either occurring naturally in clinical isolates of C. albicans as a result of natural allelic polymorphism or engineered after random mutagenesis.

7. A person skilled in the art would know that significant natural allelic variation occurs in all characterized microbial pathogens, including C. albicans (for example Miyazaki et al. (1999) Gene 236, 43-51). These natural variants contain single or multiple amino-acid changes in proteins when compared with the original reference strain, although the proteins retain biological activity as evidenced by the viability of the clinical isolates. Recognizing this, a person skilled in the art would be able to clone the ERG8 gene from any collection of clinical isolates of C. albicans using well established methods followed by the use of standard methods to determine the sequence of any one of the naturally occurring ERG8 genes, and hence the naturally occurring C. albicans ERG 8 proteins, from each clinical isolate. Comparing this sequence with the reference sequence shown in Figure 1 would rapidly identify natural variants of the ERG8 protein that, per se, will retain enzymatic activity.

8. Further, a person skilled in the art would be aware that PCR is itself mutagenic and could be used rapidly to generate multiple random variants of the C. albicans gene that could be screened for enzymatic activity. For this, such a person would design primers that would anneal to regions upstream and downstream of the ERG8 gene. These primers would be used to amplify the ERG8 gene by PCR using conditions known to

favor error-prone amplification (for example Vartanian J.P. et al. (1996) Nucleic Acids Research 24, 2627-2631). The products of the amplification would be cloned into a plasmid vector such that the gene product would be expressed in a bacterium and the activity of the resultant protein assayed using one of the methods described in the application. This would allow the rapid identification of variants of the ERG8 protein that retain enzymatic activity but which might vary from the sequence shown in Figure 1 by one or several residues.

9. I further declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true, and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

18th November 2004
Date

John David Charles Rosamond, Ph.D.

Figure 1. Alignment of C. albicans ERG8 protein (upper) with S. cerevisiae ERG8 protein (lower). A vertical line between the sequences indicates identical amino acids. Examples of highly conserved motifs are shown in bold, italicized text.

```
  3 KAFSAPGKAFLAGGYLVLEPIYDAYVTALSSRMHAVITPKGTSLKES...RIKISSPQFA 59
    |||||||| ||||||||   | | |  || |||||  | | ||  |        | ||
  5 RAFSAPGKALLAGGYLVLDTKYEAFVVGLSARMHAVAHPYG.SLQGSDKFEVRVKSKQFK 63
               .          .          .          .          .
 60 NGEWEYHISSNTE.KPREVQSRINPFLEATIFIVLAYIQPT.EAF...DLEII.IYSDPG 113
    ||| ||||      |       ||| |  |  |  |  |         |  | | ||
 64 DGEWLYHISPKSGFIPVSIGGSKNPFIEKVIANVFSYFKPNMDDYCNRNLFVIDIFSDDA 123
               .          .          .          .          .
114 YHSQEDTETKTSSNGEKTFLYHSRAITEVEKTGLGSSAGLVSVVATSLLSHFI...PNVI 170
    ||||||  |     |       ||  | || ||||||||||| |  | | | |     |
124 YHSQEDSVTE..HRGNRRLSFHSHRIEEVPKTGLGSSAGLVTVLTTALASFFVSDLENNV 181
               .          .          .          .          .
171 STNKDILHNVAQIAHCYAQKKIGSGFDVATAIYGLIVYRRFQPALINDVFQVLESDPEKF 230
           || || ||| || ||||||||| | || | |||| ||||
182 DKYREVIHNLAQVAHCQAQGKIGSGFDVAAAAYGSIRYRRFPPALISNLPDI...GSATY 238
               .          .          .          .          .
231 PTELKKLI.ESNWEFKHERCTLPYGIKLLMGDVKGGSETPKLVSRVLQWKKEKPEESSVV 289
      |  |  |  |        || |  | ||| | |||| |||  |  |      ||
239 GSKLAHLVDEEDWNITIKSNHLPSGLTLWMGDIKNGSETVKLVQKVKNWYDSHMPESLKI 298
               .          .          .          .          .
290 YDQLNSANLQFM...KELREMREKYDSDPETYIKELDHS.........VEPLTVAIKNIR 337
    |  |  ||  ||     |    |  |         |                |   ||
299 YTELDHANSRFMDGLSKLDRLHETHDDYSDQIFESLERNDCTCQKYPEITEVRDAVATIR 358
               .          .          .          .          .
338 KGLQALTQKSEVPIEPDVQTQLLDRCQEIPGCVGGVVPGAGGYDAIAVLVLENQVGNFKQ 397
          |  |   ||| ||| ||| ||   |       ||||||||||     |
359 RSFRKITKESGADIEPPVQTSLLDDCQTLKGVLTCLIPGAGGYDAIAVIT..KQDVDLRA 416
               .          .          .
398 KTLENPDYFHNVYWVDLEEQTEGVLEEK.PEDYIGL 432
     |  |   |  | | |      ||  || || | |
417 QT.ANDKRFSKVQWLDVTQADWGVRKEKDPETYLDK 451
```

Mol Gen Genet (1994) 245:716–723

ORIGINAL PAPER

Gavin Sherlock · A. Majeed Bahman
Amarbirpal Mahal · Jia-Ching Shieh · Miguel Ferreira
John Rosamond

Molecular cloning and analysis of *CDC28* and cyclin homologues from the human fungal pathogen *Candida albicans*

Received: 13 May 1994 / 10 June 1994

Abstract In the budding yeast *Saccharomyces cerevisiae*, progress of the cell cycle beyond the major control point in G1 phase, termed START, requires activation of the evolutionarily conserved Cdc28 protein kinase by direct association with G1 cyclins. We have used a conditional lethal mutation in *CDC28* of *S. cerevisiae* to clone a functional homologue from the human fungal pathogen *Candida albicans*. The protein sequence, deduced from the nucleotide sequence, is 79% identical to that of *S. cerevisiae* Cdc28 and as such is the most closely related protein yet identified. We have also isolated from *C. albicans* two genes encoding putative G1 cyclins, by their ability to rescue a conditional G1 cyclin defect in *S. cerevisiae*; one of these genes encodes a protein of 697 amino acids and is identical to the product of the previously described *CCN1* gene. The second gene codes for a protein of 465 residues, which has significant homology to *S. cerevisiae* Cln3. These data suggest that the events and regulatory mechanisms operating at START are highly conserved between these two organisms.

Key words *Candida albicans* · *CDC28* · G1 cyclins

Introduction

In budding yeast, as in all eukaryotes, the mitotic cell cycle can be divided into four intervals, G1-, S-, G2- and M phase. Overall control of cell division is achieved principally by regulating entry into S phase, the period of DNA synthesis, or into M phase when nuclear division and mitosis occur. In *Saccharomyces cerevisiae*, the major controlling event, termed START, occurs late in the G1 phase (Pringle and Hartwell 1981). At START, environmental signals such as nutrient availability or the presence of mating pheromone are monitored; only under appropriate conditions will cells traverse START and become committed to a round of mitotic division (for recent reviews see Sherlock and Rosamond 1993; Nasmyth 1993).

Passage through START requires the activation of a 34 kDa serine/threonine protein kinase, which in *S. cerevisiae* is encoded by the *CDC28* gene (Piggot et al. 1982). This protein is the functional homologue of the $cdc2^+$ gene product of the fission yeast *Schizosaccharomyces pombe* (Beach et al. 1982) and these two proteins serve as the paradigm for the *cdk* family of protein kinases in higher eukaryotes (Nurse 1990). The enzymic activity of Cdc28 at START is regulated at least in part by assembly of the kinase catalytic subunit into a complex with members of a family of labile proteins, the G1 cyclins (Richardson et al. 1989). In *S. cerevisiae*, at least nine proteins with potential G1 cyclin function have been identified and, although the roles of the different cyclins is unclear, it is thought that they may provide substrate specificity for the Cdc28 kinase complex (for example, see Cvrčková and Nasmyth 1993).

We have used *S. cerevisiae* as a surrogate genetic system to investigate the molecular mechanism of cell cycle control in the evolutionarily related yeast *Candida albicans* (Chen et al. 1984; Hendriks et al. 1989). *C. albicans* is an asexual diploid opportunistic human pathogen that is capable of growing with either a yeast or a hyphal morphology (for review see Scherer and Magee 1990). The factors that determine and regulate the morphogenetic choice seem likely to be important pathogenic determinants; although both morphologies are generally observed in disseminated infections (Odds 1987), various lines of evidence suggest a specific role for the yeast-hyphal transition in pathogenesis (Soll 1988).

Communicated by B. J. Kilbey

G. Sherlock · A. M. Bahman[1] · A. Mahal · J-C. Shieh
M. Ferreira · J. Rosamond (✉)
School of Biological Sciences, 2.205 Stopford Building,
University of Manchester, Oxford Road,
Manchester M13 9PT, UK

Present address:
[1] Department of Biochemistry, Faculty of Science,
University of Kuwait, 13059 Safat, Kuwait

As a first approach to the analysis of the *C. albicans* cell cycle and the relationship between cell cycle regulation and the yeast-hyphal dimorphic transition, we have screened a library of *C. albicans* genomic DNA for genes that rescue conditional lethal mutations in genes needed for the completion of START in *S. cerevisiae*. In this paper, we describe the isolation and molecular characterisation of *CDC28* and two cyclin homologues from *C. albicans*.

Materials and methods

Yeast strains and methods

The *S. cerevisiae* strains used in this work were: SB860 *cdc28-6 ura3-52 leu2 tyr1 trp1*; SB847 *cdc28-4 his3 leu2 ade2 ura3 metx* from Clive Price, University of Sheffield, UK; and BF305-15dno.21 *MATa leu2-3,-112 his3 ura3 trp1 ade1 met14 arg5,6 HIS3::cln1 TRP1::cln2 ura3::GAL1-CLN3* from Bruce Futcher, Cold Spring Harbor Laboratory, New York (Xiong et al. 1991). *C. albicans* strain 124 was obtained from Richard Barton, University of Manchester. All strains were grown on media containing 2% peptone, 1% yeast extract supplemented with either 2% glucose (YEPD) or 1% galactose and 1% raffinose (YEPGR). Supplemented synthetic minimal medium (YNB) comprising 0.67% yeast nitrogen base, 2% glucose and appropriate nutritional supplements was used for the selection and maintenance of plasmids in *S. cerevisiae*. Standard yeast genetic and recombinant techniques were used (Sherman et al. 1986). Yeasts were transformed using the lithium acetate procedure with single-stranded carrier DNA (Schiestl and Gietz 1989).

Bacterial strains and methods

Escherichia coli strain HW87 (Patterson et al. 1986) was used as the routine host for maintenance and storage of plasmids. Cultures were typically grown in L-broth (Miller 1972) supplemented when necessary with 50 μg/ml ampicillin. Plasmid DNA was extracted from bacterial cultures either by alkaline lysis (Birnboim and Doly 1979) or by detergent lysis followed by CsCl-ethidium bromide equilibrium density gradient centrifugation (Humphreys et al. 1975). *E. coli* HW87 was transformed either by the method of Warren and Sherratt (1978) or by electroporation (Dower et al. 1988).

Nucleic acid methods

Yeast genomic DNA was prepared from logarithmic-phase cell cultures as described previously (Cryer et al. 1975). Standard recombinant DNA techniques were used throughout (Sambrook et al. 1989). Restriction endonucleases, T4 DNA ligase and Klenow fragment of DNA polymerase I were purchased from Boehringer Mannheim. Sequenase version 2.0 was purchased from United States Biochemical Co. and used according to the manufacturer's recommendations. DNA fragments were radioactively labelled by random oligonucleotide priming (Feinberg and Vogelstein 1983) using a kit from Boehringer Mannheim and [α-^{32}P]dATP from New England Nuclear. DNA sequences were determined using the chain-termination method (Sanger et al. 1977) for direct plasmid sequencing on both strands (Zhang et al. 1988) using [α-^{35}S]dATP. Oligonucleotide primers were synthesised on an ABI381 Synthesiser using phosphoramidite chemistry. Reaction products were resolved and detected as described previously (Patterson et al. 1986). The deduced sequence was analysed using University of Wisconsin Genetics Computer Group (GCG) software on the Daresbury database facility.

Results

Construction of a *C. albicans* genomic library

A high-copy-number library of genomic DNA sequences from *C. albicans* strain 124 was generated using high molecular weight DNA that had been partially digested with *Sau*3A. The digested DNA was size fractionated to 3–12 kb by centrifugation through 5–20% neutral sucrose gradients (Rosamond et al. 1979) and cloned into the *Bam*HI site of the shuttle vector YEp24, which carries the yeast *URA3* gene and 2 μm replication origin (Hurley and Donelson 1980). The library contains 2.5×10^4 independent plasmids of which 70% are recombinant. The average size of the inserts is 7.3 kb. Since YEp24 lacks sequences to direct the expression of cloned DNA, the expression of genes from within the *C. albicans* genomic DNA inserts of the recombinant plasmids relies upon adjacent *C. albicans* regulatory elements.

Cloning and identification of *CaCDC28*

The *CaCDC28* gene was cloned by screening the *C. albicans* genomic library in YEp24 for genes that could suppress the temperature-sensitive lesion in *S. cerevisiae* SB860 (*cdc28-6*). Approximately 6000 Ura$^+$ transformants were obtained initially at 23°C. These cells were recovered in pools of approximately 10^3 transformants and aliquots of each pool were replated on YEPD agar at 37°C. Plasmid DNA was rescued from colonies that grew at the restrictive temperature, amplified in *E. coli* and used to re-transform *S. cerevisiae* SB860 to uracil prototrophy and temperature resistance. From this screen, a single plasmid was isolated that carried a 7.5 kb genomic fragment capable of rescuing both the *S. cerevisiae cdc28-6* and *cdc28-4* mutations. This plasmid was designated pANJ1 (Fig. 1A).

To delimit the region of the genomic fragment cloned in pANJ1 that was responsible for complementation of *cdc28*, we subcloned portions of pANJ1 and tested the ability of the subclones to rescue growth at the restrictive temperature in *S. cerevisiae* SB860. A subclone carrying the 3.5 kb *Sph*I fragment (pANJ2; Fig. 1A) was unable to complement *cdc28-6*. However, a subclone that carried the 3.8 kb region from the *Sph*I site to the end of the insert (pANJ3) was able to rescue *cdc28* as effectively as pANJ1 (Fig. 1A). We conclude therefore that pANJ3 contains all of the elements essential for complementation of *cdc28*.

Nucleotide sequence of *CaCDC28*

Using synthetic oligonucleotide primers, we have determined the complete nucleotide sequence of *CaCDC28* within the cloned DNA fragment of pANJ3. The se-

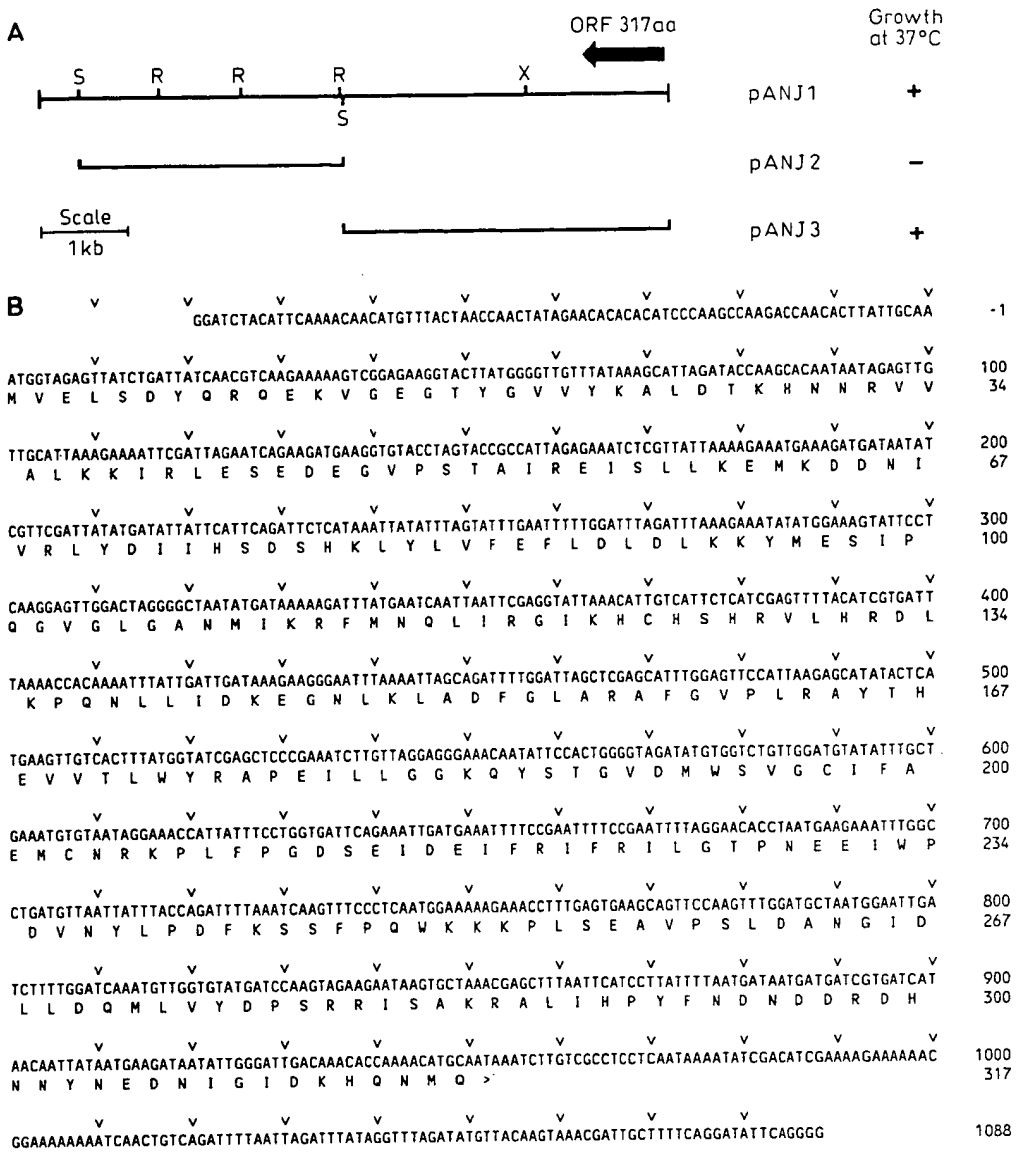

Fig. 1A,B Characterisation of *CaCDC28*. **A** Partial restriction map and complementation analysis of *CaCDC28* subclones. Complementation was assayed by the ability of subclones to restore growth of the *cdc*28-6 strain at 37°C; + indicates growth, − indicates no growth. The *large arrow* shows the location, size and direction of transcription within the cloned DNA of the *CaCDC28* open reading frame. Abbreviations of restriction enzyme sites are as follows: R, *Eco*RI; S, *Sph*I; X, *Xho*I. **B** Nucleotide and deduced amino acid sequence of *Candida albicans CDC28*. Nucleotides are numbered with respect to the first ATG of the open reading frame (ORF). This nucleotide sequence will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X80034

quence contains an open reading frame of 317 codons, potentially encoding a protein of 36645 daltons, the location of which is consistent with the subcloning data (Fig. 1A, B). The regions flanking the open reading frame contain motifs frequently found adjacent to coding regions in yeast, including a TATA box at position −40 relative to the initiation codon, as well as consensus transcription termination and polyadenylation signals in the 3′ flanking region between nucleotides 971–1073 (Fig. 1B; Zaret and Sherman 1982). Comparison of the predicted protein product of *CaCDC28* with *S. cerevisiae* Cdc28 showed that the proteins were 79% identical over 295 amino acids, and that CaCdc28 contains all of the motifs characteristic of protein kinases in general and the Cdc28 protein kinase in particular (Fig. 2; Hanks and Quinn 1991).

Cloning and identification of *CaCLN* genes

Since Cdc28 protein kinase activity is regulated in part by interaction with cyclins (Richardson et al. 1989), we screened the *C. albicans* genomic library for genes encoding putative G1 cyclins. For this purpose we used *S. cerevisiae* strain BF305-15dno.21, which is deleted for *CLN1* and *CLN2* and dependent for viability on the galactose-inducible expression of *CLN3* (Xiong et al. 1991). *S. cerevisiae* BF305-15dno.21 was grown in YEP-GR and transformed with DNA from the *C. albicans* genomic library. Cells were screened for plasmid-borne cyclin genes by plating directly onto minimal YNB medium lacking uracil and supplemented with glucose. Two colonies were obtained (from a total of approximately 5×10^3 Ura^+ transformants); plasmid DNA was recovered from each of these clones, amplified in *E. coli* and used to re-transform BF305-15dno.21 to uracil prototrophy and galactose independence. From this screen

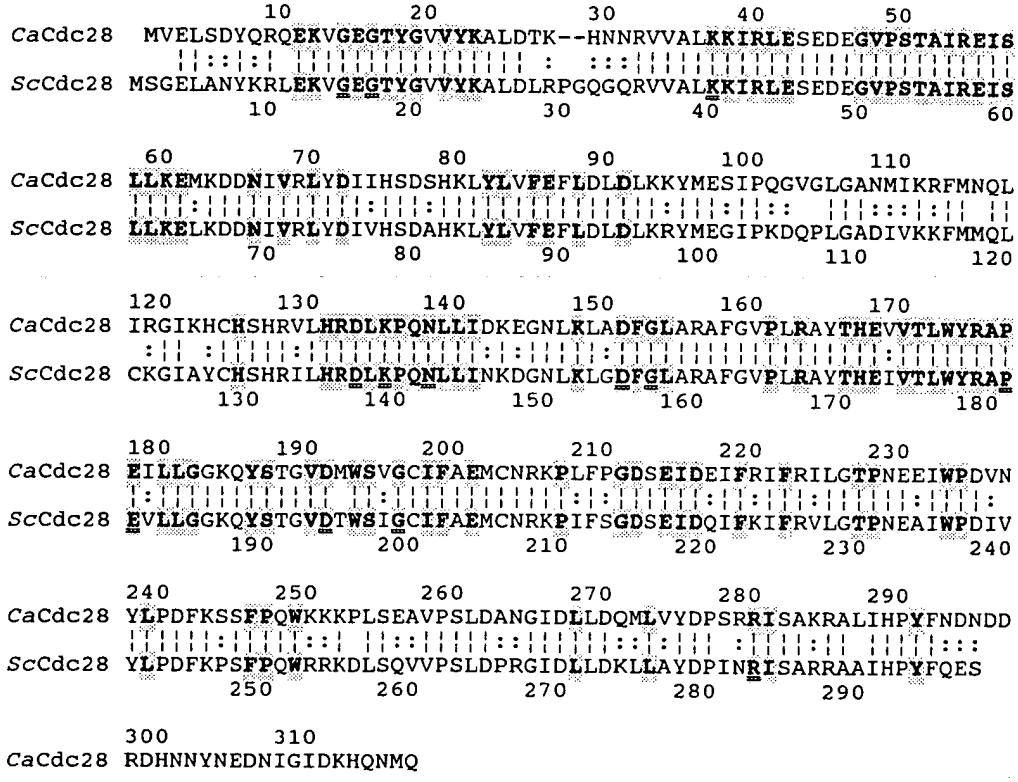

Fig. 2 Comparison of the *Candida albicans* and *Saccharomyces cerevisiae* Cdc28 proteins. Proteins were aligned using FASTA on the Daresbury SEQNET facility. Amino acid identities are indicated by *dashes*; conservative substitutions are indicated by *colons*. Key residues conserved within protein kinases are *underlined*; residues characteristic of Cdc28/cdc2 kinases are *shaded*.

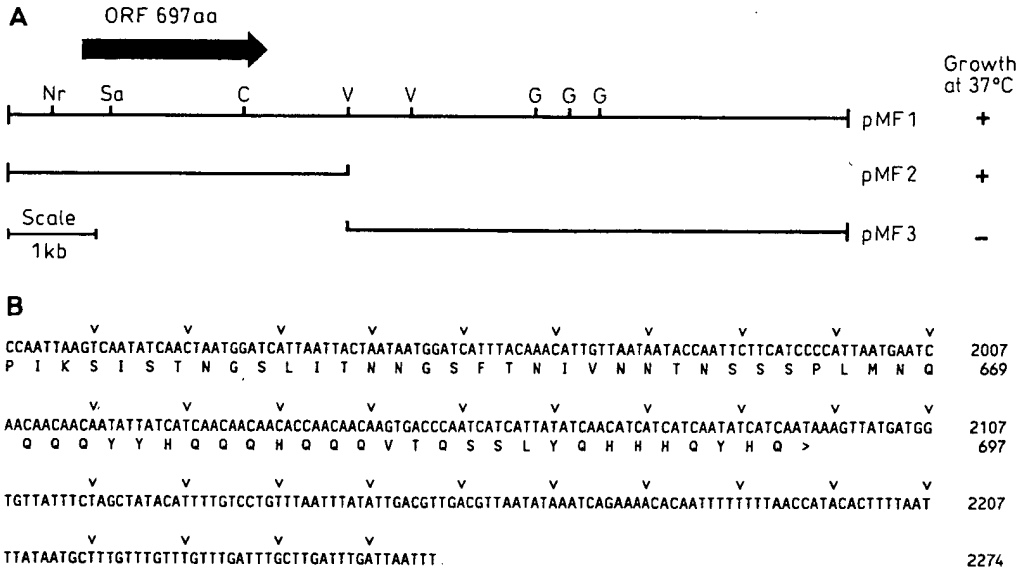

Fig. 3A,B Characterisation of *CaCLN1*. **A** Partial restriction map and complementation analysis of *CaCLN1* subclones, for which symbols used are as described in the legend to Fig. 1. Restriction site abbreviations: C, *Cla*I; G, *Bgl*II; Nr, *Nru*I; Sa, *Sal*I; V, *Eco*RV. **B** Nucleotide sequence of the C-terminal region and 3′-flanking region of *CaCLN1*. This sequence, together with 616 bp of 5′-flanking sequence, will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X80032 as an update to the previously reported partial ORF sequence (accession number M76587; Whiteway et al. 1992)

we obtained two independent plasmids, designated pMF1 and pGS1, which were defined as carrying *CaCLN1* and *CaCLN2*, respectively.

The genomic fragment cloned in pMF1, carrying *CaCLN1*, was characterised to localise the region required for complementation. This was achieved by subcloning portions of pMF1 and examining the ability of the subclones to rescue the conditional cyclin deficiency in *S. cerevisiae* BF305-15dno.21. The results (Fig. 3A) established that all of the elements required for complementation were located in the 4.0 kb region extending from an *Eco*RV site to the left vector-insert junction (pMF2; Fig. 3A). In the course of this work, it became apparent that the clone carrying *CaCLN1* was similar to the clone carrying the previously reported partial open reading frame, *CaCCN1*, which had been isolated on the basis of its ability to confer α-factor resistance in *S. cerevisiae* (Whiteway et al. 1992). From DNA sequencing we have established that *CaCLN1* is identical to *CaCCN1* for which a partial amino acid sequence has

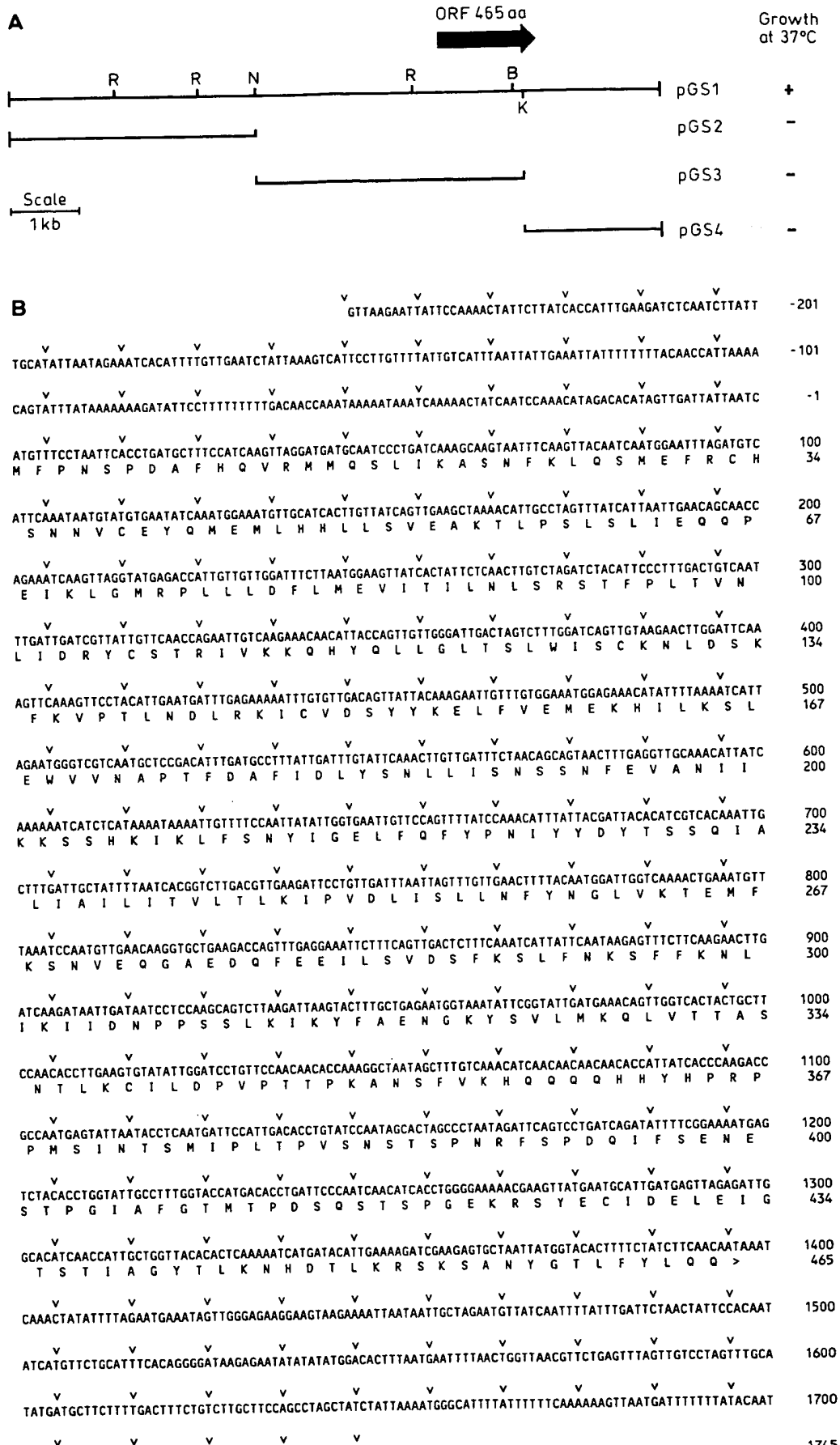

Fig. 4A,B Characterisation of *CaCLN2*. **A** Partial restriction map and complementation analysis of *CaCLN2* subclones, symbols used are as described in the legend to Fig. 1. Restriction enzyme abbreviations: B, *Bam*HI; K, *Kpn*I; N, *Nhe*I; R, *Eco*RI. The *large arrow* shows the location, size and direction of transcription of the *CaCLN2* open reading frame within the cloned DNA. **B** Nucleotide and deduced amino acid sequence of *C. albicans CLN2*. Nucleotides are numbered with respect to the first ATG of the open reading frame. Regulatory features are described in the text. The nucleotide sequence given here will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X80033

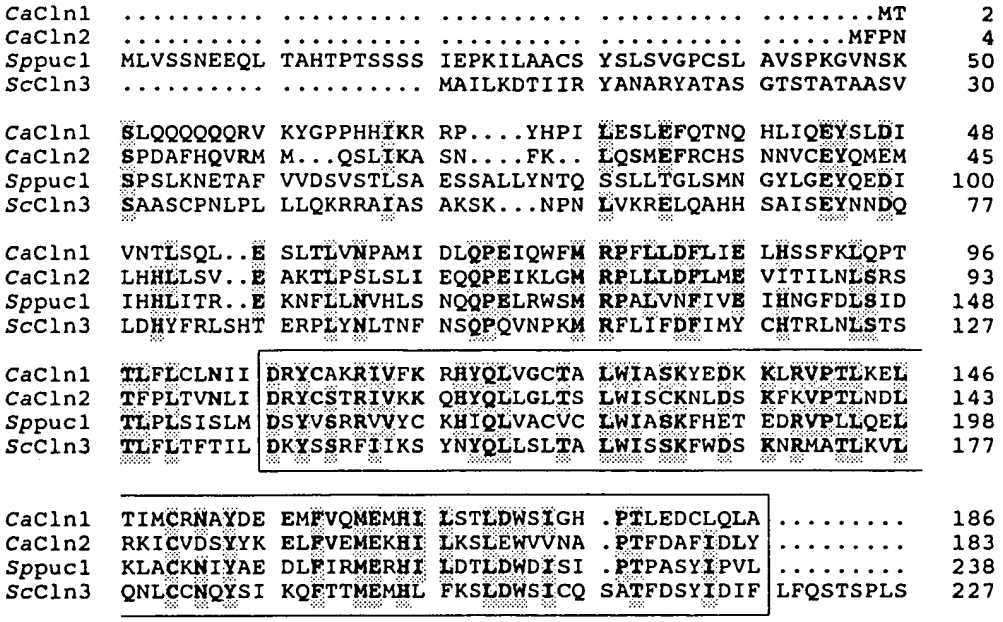

Fig. 5 Comparison of the N-termini of *Ca*Cln1, *Ca*Cln2, *S. cerevisiae* Cln3 (*Sc*Cln3) and *Schizosaccharomyces pombe* puc1$^+$ (*Sp*puc1). Sequences were aligned using the University of Wisconsin Genetics Computer Group (GCG) PILEUP program. *Dots* represent gaps introduced to maximise the alignment. Residues conserved in at least three of the proteins are *shaded*, while amino acids conserved between *Ca*Cln1 and *Ca*Cln2 are shown in *bold*. The cyclin box is *outlined*

been reported (Whiteway et al. 1992). We have confirmed that sequence and extended it to complete the sequence of the entire open reading frame together with 616 bp of 5′-flanking sequence and 158 bp of downstream sequence. Figure 3B shows the nucleotide and predicted protein sequence of the previously unreported C-terminal portion of CaCln1. The predicted gene product thus comprises 697 amino acids with a molecular mass of 79250 daltons.

Characterisation and nucleotide sequence of *CaCLN2*

The genomic fragment in pGS1 was characterised by restriction mapping, which demonstrated that it was unrelated to CaCLN1, and by subcloning to localise the complementing element. All three subclones derived from pGS1 failed to complement the conditional cyclin deficiency in strain BF305-15dno.21 (pGS2-4; Fig. 4A) implying that *CaCLN2* spanned either the *NheI* or *KpnI* site, or both. Using synthetic oligonucleotide primers,initial sequencing experiments demonstrated that the *KpnI* site was contained within an open reading frame, the predicted product of which had significant homology to *S. cerevisiae* Cln3. We have completed the sequence of the *CaCLN2* gene together with 255 nucleotides of 5′- and 347 nucleotides of 3′- flanking sequence (Fig. 4B). The gene encodes a protein of 465 amino acids with a molecular mass of 52000 daltons. In addition, the gene is flanked by typical regulatory motifs including a TATA box at position −90 relative to the initiation codon, and transcription termination and polyadenylation signals in the 3′-flanking region between nucleotides 1670–1750 (Fig. 4).

Comparison of CaCln2 with the EMBL protein sequence database revealed that the protein was most similar to the *S. cerevisiae* G1 cyclin Cln3, *C. albicans* Ccn1 (CaCln1) and *S. pombe* puc1$^+$, all of which complement the conditional G1 cyclin mutation in BF305-15dno.21. This homology is most obvious within the so-called cyclin box region of the proteins found in the N-terminal region of both CaCln1 and CaCln2 (Fig. 5). Within this domain, the proteins show 36–40% identity in pairwise comparisons. In addition, and in contrast to the parental strain which is unable to grow in the presence of 1 μM α-factor, cells overexpressing either *CaCLN1* or *CaCLN2* are resistant to mating pheromone and grow well in media supplemented with 10 μM α-factor. In both cases, overexpression occurs only as a consequence of the elevated copy number of the gene in YEp24, since both *C. albicans* genes are expressed in *S. cerevisiae* from their native promoters. In *S. cerevisiae*, the pheromone response pathway inhibits G1 cyclin activity to bring about cell cycle arrest (for review see Kurjan 1993). Therefore, while the finding is not altogether surprising that the overexpression of *CaCLN1* and *CaCLN2* results in α-factor resistance, the significance of this with respect to a possible response pathway in *C. albicans* is not clear.

Discussion

We have described the cloning and molecular characterisation of *CDC28* and two putative G1 cyclins from the human fungal pathogen *C. albicans*. The *CDC28* gene from *C. albicans* rescues *cdc28-6* and *cdc28-4* mutations in *S. cerevisiae*, while the protein products of the *C. albicans* and *S. cerevisiae* genes share 79% identity, making these two proteins the most similar within the *cdk* family (Hanks and Quinn 1991). The only difference worthy of note is a short extension of 22 residues at the C-terminus

of *C. albicans* Cdc28, which distinguishes it from other members of the p34 family though the functional significance of this difference is unclear.

In addition to *CaCDC28*, we have isolated two genes that encode cyclins in *Candida*. On the basis of the finding that the expression of these genes can rescue a triple *cln* mutation in *S. cerevisiae*, we suggest that these proteins have a similar function in *C. albicans*, and act as G1 cyclins to regulate Cdc28 protein kinase activity at START. This idea is supported by the observation that, in addition to being homologous to one another, both *Candida* cyclins are most similar to *S. cerevisiae CLN3* and *S. pombe puc1*$^+$, both of which also rescue a triple *cln* mutant (Forsburg and Nurse 1991). However, *puc1*$^+$ now appears to function as a meiotic rather than a G1 cyclin in *S. pombe* (Forsburg and Nurse 1994), and since *Candida* is unable to undergo meiosis, the significance of this similarity is unclear. *CLN1*, *CLN2* and *CLN3* form a functionally redundant gene family in *S. cerevisiae* (Richardson et al. 1989), though there are clear differences in the regulation and function of each of the gene products.

CLN3 is transcribed constitutively and regulated post-translationally by Swi4 and Swi6, whereas the transcription of both *CLN1* and *CLN2* is periodically regulated, peaking in G1 phase, by the Swi4/Swi6 transcription factor, SBF (Nasmyth and Dirick 1991). Also, it has been suggested that Cln3 functions upstream of Cln1 and Cln2 in order to regulate their activity (Tyers et al. 1993). We find that *CaCLN1* has sequences in its 5′-flanking region that are consistent with cell cycle regulation of its transcript in a manner analogous to *CLN1* and *CLN2* in *S. cerevisiae* (G. Sherlock, A. M. Bahman and J. Rosamond, in preparation). However, we have found no such sequence motifs upstream of *CaCLN2*, suggesting that it may in fact have a function more related to that of Cln3 than to Cln1 or Cln2. Such a categorisation requires, however, an analysis of the pattern of expression of these genes during growth and development; such an analysis is currently in progress.

Acknowledgements We would like to thank Clive Price and Bruce Futcher for gifts of plasmids and strains; Anne White, Mark Buddles, Eliot Randle and Penny Johnson for advice, assistance and useful comments. We gratefully acknowledge financial assistance from SERC (to J. R.) and the award of a Wellcome Prize Studentship (to G. S.).

References

Beach D, Durkacz B, Nurse P (1982) Functionally homologous cell cycle genes in budding and fission yeast. Nature 300:706–709

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acid Res 7:1513–1523

Chen M-W, Anne J, Volckaert G, Huysmans E, Vandenberghe A, Wachter RD (1984) The nucleotide sequence of the 5S rRNA of seven molds and a yeast and their use in studying Ascomycete phylogeny. Nucleic Acids Res 12:4881–4892

Cryer DR, Eccleshall R, Marmur J (1975) Isolation of yeast DNA. Methods Cell Biol 12:39–44

Cvrčková F, Nasmyth K (1993) Yeast G1 cyclins *CLN1* and *CLN2* and a GAP-like protein have a role in bud formation. EMBO J 13:5277–5286

Dower, WJ, Miller, JF, Ragsdale, CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16:6127–6145

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Forsburg SL, Nurse P (1991) Identification of a G1-type cyclin *puc1*$^+$ in the fission yeast *Schizosaccharomyces pombe*. Nature 351:245–248

Forsburg SL, Nurse P (1994) Analysis of the *Schizosaccharomyces pombe* cyclin puc1: evidence for a role in cell cycle exit. J Cell Sci 107:601–613

Hanks SK, Quinn AM (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. In: Hunter T, Sefton BM (eds) Methods Enzymol 200:38–62

Hendriks L, Goris A, Neffs J-M, Peer YVD, Hennebert G, Wachter RD (1989) The nucleotide sequence of the small ribosomal subunit RNA of the yeast *Candida albicans* and the evolutionary position of the fungi among eukaryotes. Syst Appl Micribiol 13:223–229

Humphreys GO, Willshaw GA, Anderson ES (1975) A simple method for the preparation of large quantities of pure plasmid DNA. Biochim Biophys Acta 383:457–463

Hurley JL, Donelson JE (1980) Nucleotide sequence of the yeast plasmid. Nature 286:860–865

Kurjan J (1993) The pheromone response pathway in *Saccharomyces cerevisiae*. Annu Rev Genet 27:147–179

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Nasmyth K (1993) Control of the yeast cell cycle by the Cdc28 protein kinase. Curr Biol 5:166–179

Nasmyth K, Dirick L (1991) The role of SWI4 and SWI6 in the activity of G1 cyclins in yeast. Cell 66:995–1013

Nurse P (1990) Universal control mechanism regulating onset of M-phase. Nature 344:503–508

Odds FC (1987) Candida infection: an overview. Crit Rev Microbiol 15:1–5

Patterson MN, Sclafani RA, Fangman WL, Rosamond J (1986) Molecular characterization of cell cycle gene *CDC7* from *Saccharomyces cerevisiae*. Mol Cell Biol 6:1590–1598

Piggot JR, Rai R, Carter BLA (1982) A bifunctional gene product involved in two phases of the yeast cell cycle. Nature 298:391–393

Pringle JR, Hartwell LH (1981) The *Saccharomyces cerevisiae* cell cycle. In: Strathern JN, Jones EW, Broach JR (eds) The life cycle of the yeast *Saccharomyces cerevisiae*: life cycle and inheritance, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York pp 97–142

Richardson HE, Wittenburg C, Cross F, Reed SI (1989) An essential G1 function for cyclin-like proteins in yeast. Cell 59:1127–1133

Rosamond J, Telander KM, Linn S (1979) Modulation of the action of the *recBC* enzyme of *Escherichia coli* K-12 by Ca^{2+}. J Biol Chem 254:8646–8652

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Scherer S, Magee PT (1990) Genetics of *Candida albicans*. Microbiol Rev 54:226–241

Schiestl RH, Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr Genet 16:339–346

Sherlock G, Rosamond J (1993) Starting to cycle: G1 controls regulating cell division in budding yeast. J Gen Microbiol 139:2531–2541

Sherman F, Fink GR, Hicks (1986) Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Soll DR (1988) High frequency switching in *Candida albicans* and its relations to vaginal candidiasis. Am J Obstet Gynecol 154:997–1001

Tyers M, Tokiwa G, Futcher B (1993) Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. EMBO J 12:1955–1968

Warren G, Sherratt D (1978) Incompatibility and transformation efficiency of ColE1 and related plasmids. Mol Gen Genet 161:39–47

Whiteway M, Dignard D, Thomas DY (1992) Dominant negative selection of heterologous genes: isolation of *Candida albicans* genes that interfere with *Saccharomyces cerevisiae* mating factor-induced cell cycle arrest. Proc Natl Acad Sci USA 89:9410–9414

Xiong Y, Connolly T, Futcher B, Beach D (1991) Human D-type cyclin. Cell 65:691–699

Zaret K, Sherman F (1982) DNA sequence required for efficient transcription termination in yeast. Cell 28:563–573

Zhang H, Scholl R, Browse J, Somerville C (1988) Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res 16:1220

ELSEVIER

Gene 183 (1996) 159–165

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Isolation and molecular characterisation of the *POL3* gene from *Candida albicans*

Tania Nolan, John Rosamond *

School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester, M13 9PT, UK

Received 14 March 1996; revised 17 June 1996; accepted 19 June 1996

Abstract

In *Saccharomyces cerevisiae*, the *CDC2* gene encodes the large subunit of DNA polymerase III, the analogue of mammalian DNA polymerase δ. We have isolated DNA fragments from a library of *Candida albicans* genomic DNA in the vector pRS316 that rescue temperature sensitive *cdc2* mutations in *S. cerevisiae*. These fragments contain an ORF coding for a protein of 1038 aa with a predicted molecular mass of 118.8 kDa. The predicted protein shows homology to a number of eukaryotic DNA polymerases, with 62% identity over its length to the *S. cerevisiae* Cdc2 protein. It also contains a number of motifs which are characteristic of DNA polymerases in general and viral polymerases in particular, as well as the conserved motif which interacts with proliferating cell nuclear antigen. These results indicate that this gene is *C. albicans POL3*. Analysis of the expression of *C. albicans POL3* revealed that the transcript is present throughout the mitotic cell cycle, which contrasts with the expression of *S. cerevisiae CDC2*.

Keywords: *Candida albicans*; DNA replication; DNA polymerase; *POL3*; Gene expression; Cell cycle

1. Introduction

Control of cell division in *Saccharomyces cerevisiae* is achieved principally at a point in G1 termed START, completion of which commits the cell to a round of division and sets in train the events required for the initiation of DNA synthesis. Progress through START is governed chiefly by modulation of the activity of the Cdc28 protein kinase which is achieved by two mechanisms: one entails the degradation of an inhibitor of Cdc28 kinase activity in a ubiquitin-dependent reaction (Schwob et al., 1994); the other requires activation by interaction with a number of labile proteins, termed G1 cyclins (Forsburg and Nurse, 1991; Nasmyth, 1993). One of these complexes, Cln3-Cdc28, activates two transcription factors: SBF which comprises Swi4 and Swi6 proteins and MBF (or DSC1), which is composed of the Mbp1 and Swi6 proteins (Tyers et al., 1993; Koch et al., 1993; reviewed in Sherlock and Rosamond, 1993; McIntosh, 1993). MBF binds to a specific sequence within the promoter of over 20 genes whose products are components of the enzymatic machinery for DNA synthesis as well as enzymes involved in the biosynthesis of precursors, thereby coordinating the expression of these genes within late G1 and early S phases of the cell cycle (McIntosh, 1993). Amongst the genes which are regulated in this way in *S. cerevisiae* is *CDC2*, which encodes the large subunit of DNA polymerase III, the analogue of mammalian DNA polymerase δ (Boulet et al., 1989).

We are interested in the enzymology and control of DNA replication in the human fungal pathogen *Candida albicans*, and in particular the potential use of components of the replication machinery as targets for novel anti-fungal drugs. In contrast to our understanding of the DNA replication machinery and its regulation in *S. cerevisiae*, relatively little is known of the equivalent processes in the related dimorphic yeast *C. albicans*, in part because the organism is an obligate diploid that lacks a sexual cycle (Odds, 1988) and uses a non-standard genetic code (Santos and Tuite, 1995). Previous work has identified DNA polymerase activity in extracts of *C. albicans*. Surprisingly though, none of the activities detected had properties consistent with DNA polymerase III (Jakab et al., 1991).

* Corresponding author. Tel. +44 161 2755100; Fax +44 161 2755082; e-mail: rosamond @fs2.scg.man.ac.uk

Abbreviations: aa, amino acid(s); bp, base pair(s); kb, kilobase(s) or 1000 bp; kDa, kilodalton; nt, nucleotide(s); ORF, open reading frame.

PII S0378-1119(96)00549-5

We have used a surrogate genetic approach to identify genes from *C. albicans* that encode proteins necessary for DNA replication and in particular to identify the gene encoding DNA polymerase III. In this paper we describe the isolation and molecular characterisation of a gene encoding a DNA polymerase III homologue from *C. albicans*, designated *POL3*, by complementation of a conditional mutation in *S. cerevisiae CDC2*. This is the first gene encoding a component of the replication machinery to be isolated from this organism. Additionally we demonstrate that, in contrast to the periodic expression of *CDC2* in *S. cerevisiae*, there is little variation in the expression of *POL3* during the budding cell cycle of *C. albicans*.

2. Experimental and discussion

2.1. *Identification of* C. albicans *genomic DNA fragments which suppress the* S. cerevisiae cdc2-2 *mutation*

Genomic DNA fragments capable of complementing the temperature sensitive *cdc2-2* mutation in *S. cerevisiae* were isolated from a library of random *Sau*3A fragments of *C. albicans* DNA in the vector pRS316 (Sikorski and Hieter, 1989; Sherlock et al., 1994). pRS316 lacks sequences which direct the expression of cloned DNA and consequently the expression of genes from the *C. albicans* genomic DNA inserts requires adjacent *C. albicans* regulatory elements.

S. cerevisiae strain SB667 (*ura3 cdc2-2*) was transformed with the *C. albicans* library DNA and *URA3*$^+$ transformants were selected by growth of colonies at 23°C for 5 days on agar medium lacking uracil. Approximately 1×10^5 independent transformants were recovered in pools of about 5×10^3 clones, replated onto YPD agar and incubated at 34°C for 3 days. Plasmid DNA was recovered from representative colonies which grew at 34°C and amplified through *E. coli* HW87. Individual plasmid isolates were then used to re-transform *S. cerevisiae* strain SB667 to uracil prototrophy and temperature resistance. In this way, we isolated two independent plasmids that were capable of rescuing the *cdc2-2* mutation at 34°C although neither plasmid could restore viability at 37°C; these plasmids were designated pTAN1 and pTAN2.

2.2. *Characterisation of the cloned* C. albicans *genomic fragments*

Restriction mapping showed that the genomic inserts of pTAN1 and pTAN2 were different but related and that pTAN1 contained a 4.0-kb genomic DNA fragment while pTAN2 contained a 6.2-kb genomic DNA fragment (Fig. 1). This analysis established that the sequence cloned in pTAN1 was contained entirely within pTAN2 while Southern blotting demonstrated that the cloned DNA fragment was present as a single copy in the *Candida* genome (data not shown). Since, on the basis of DNA polymerase III-like proteins in other organisms, we would expect the *POL3* coding and regulatory sequences to extend for about 3.5 kb, pTAN1 was used for further investigation without subcloning.

We have determined the nucleotide sequence of the complete *C. albicans* genomic fragment cloned in pTAN1. This sequence contains a single significant ORF capable of encoding a protein of 1038 aa. The nucleotide sequence of this ORF is shown in Fig. 2, together with the aa sequence of the predicted protein, 550 nucleotides of 5′ flanking sequence and 60 nucleotides of 3′ flanking sequence. The ORF encodes a putative 118-kDa protein that is 58.9% identical to polδ$^+$, the DNA polymerase III homologue from *S. pombe* (Park et al., 1993) and 62.6% identical to the Cdc2 protein sequence from *S. cerevisiae* (Boulet et al., 1989) over a region of 993 aa. The predicted product of the ORF also contains conserved motifs including: (a) six regions characteristic of DNA polymerases in the same spatial array as *S. cerevisiae* Cdc2 (Fig. 3A; Wong et al., 1988); (b) four regions that define a viral sub-family of polymerases (Fig. 3B; Kouzarides et al., 1987); (c) two potential zinc-finger DNA-binding domains (Fig. 2); (d) a motif between aa 83 and 103 that is the site of interaction for DNA polymerase III-type enzymes and proliferating cell nuclear antigen (Zhang et al., 1995). From these data, we conclude that this ORF encodes *C. albicans* DNA polymerase III and accordingly have designated the gene *POL3*.

2.3. *Analysis of the* C. albicans POL3 *transcript levels during the cell cycle*

S. cerevisiae CDC2 RNA accumulates periodically in *S. cerevisiae* in response to MBF transcription factor, for which there is a single binding site 140 bp upstream of the initiation codon (Boulet et al., 1989; Bauer and Burgers, 1990). Examination of the *C. albicans POL3* promoter revealed a possible MBF site 63 bp upstream of the translation start site (Fig. 2). To examine whether *C. albicans* might have a regulatory system analogous to MBF and whether *POL3* might be periodically regulated by such a system, we prepared RNA from aliquots of a *C. albicans* culture undergoing synchronous division, immobilised the RNA onto a Hybond-N filter and probed for *C. albicans CDC2* and *S. cerevisiae* actin. The results of a single representative experiment are shown in Fig. 4; similar results have been obtained in other independent experiments.

As observed previously (Mitchell and Soll, 1979) on release from arrest, there is a delay of approximately 2 h before the first appearance of small buds. The

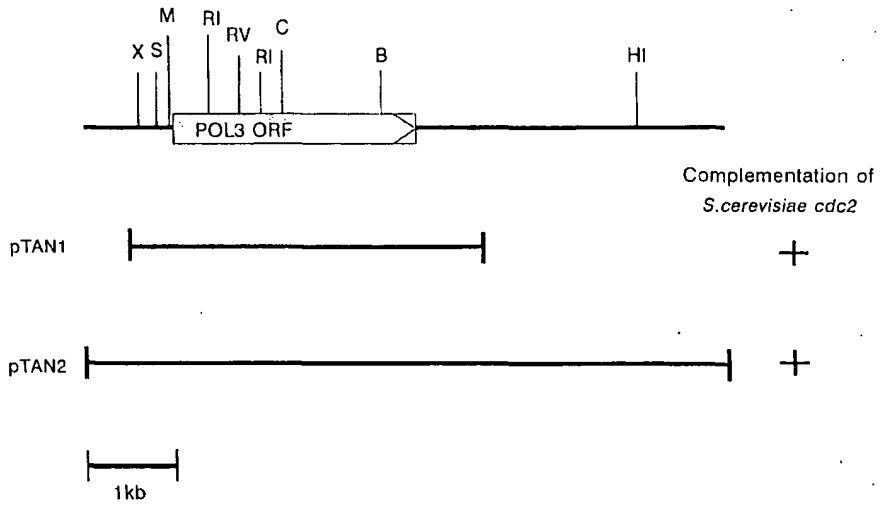

Fig. 1. Restriction map of the *cdc2*-complementing clones isolated from a *C. albicans* genomic library. The upper line shows a composite map of the locus, with the location of the *C. albicans POL3* open reading frame indicated. Restriction enzymes are abbreviated as; X, *Xho*I; S, *Spe*I; M, *Mlu*I; RI, *Eco*RI; RV, *Eco*RV; C, *Cla*I;B, *Bst*XI; HI, *Bam*HI. *S. cerevisiae* strain SB667 (MATa *cdc2-2, ura3-52 trp1 lys2*) was transformed with a library of *C. albicans* genomic DNA in pRS316 (Sikorski and Hieter, 1989) using alkali cations (Schiestl and Gietz, 1989; Gietz et al., 1992) and plasmids capable of restoring viability at 34°C were selected and purified through *E. coli* HW87 (Patterson et al., 1986). Maintenance, preparation and manipulation of plasmid DNA followed standard protocols (Sherman et al., 1986; Dower et al., 1988; Sambrook et al., 1989).

number of cells with small buds reaches a maximum around 220 min after release from arrest (Fig. 4a), approximately coincident with the initiation of DNA synthesis in S phase, before declining as the cells proceed through the cell cycle. We have confirmed the physiological state of the cells by FACS analysis which showed that arrested cells are exclusively in G1 phase, while at 220 and 300 min post-arrest, the cells are in S and G2 phases respectively (Fig. 4a). Significantly, *C. albicans POL3* transcript is detectable throughout this period, with changes in signal intensity largely paralleling variations in total RNA as reflected by the ethidium bromide-stained gel (Fig. 4b and c). When compared with actin, the *POL3* transcript fluctuates slightly through the cell cycle, with maximum *POL3* mRNA levels coinciding with the onset of S phase (Fig. 4a). However, once cells have recovered from the stress of arrest, this fluctuation is small (of the order of 2 to 3-fold) and suggests that *C. albicans POL3* is not periodically regulated in a manner analogous to *CDC2* in *S. cerevisiae*.

3. Conclusions

(1) We have described the isolation of two clones from a library of *C. albicans* genomic DNA which can restore viability at 34°C in a strain of *S. cerevisiae* carrying the thermosensitive *cdc2-2* mutation. These clones are related in that the smaller 4 kb cloned fragment (pTAN1, Fig. 1) is contained entirely within the genomic sequences of the larger fragment. The 4-kb fragment of pTAN1 was sequenced completely and shown to encode *C. albicans POL3*.

(2) pTAN1 will be maintained in *S. cerevisiae* at approximately 1–2 copies per cell, since the vector on which this plasmid is based contains a functional centromere (Sikorski and Hieter, 1989). In addition, *POL3* is being expressed from the normal cognate *C. albicans* flanking sequences rather than from a *S. cerevisiae* promoter. This suggests firstly that the gene is not highly expressed, consistent with the low codon bias index of both *CDC2* and *POL3*, and also that the level of expression is not critical to the function of DNA polymerase III. It is also consistent with the view that not only do the *S. cerevisiae* and *C. albicans* proteins display a high degree of primary sequence homology, but that other aspects of their activity, such as kinetic parameters and their ability to interact productively with accessory proteins within the replication complex such as PCNA (Bauer and Burgers, 1990) are also similar.

(3) Complementation of *cdc2* by *POL3* is clearly incomplete as cells carrying pTAN1 fail to grow at 37°C and arrest at this temperature with the dumbbell morphology characteristic of *cdc2* strains (Culotti and Hartwell, 1971). This may reflect a difference in the regulation of DNA polymerase III expression in the two organisms. In *S. cerevisiae, CDC2* is regulated periodically during the cell cycle, with maximum expression observed during late G1 and early S phases (Bauer and Burgers, 1990). This is probably mediated by the MBF transcription factor,

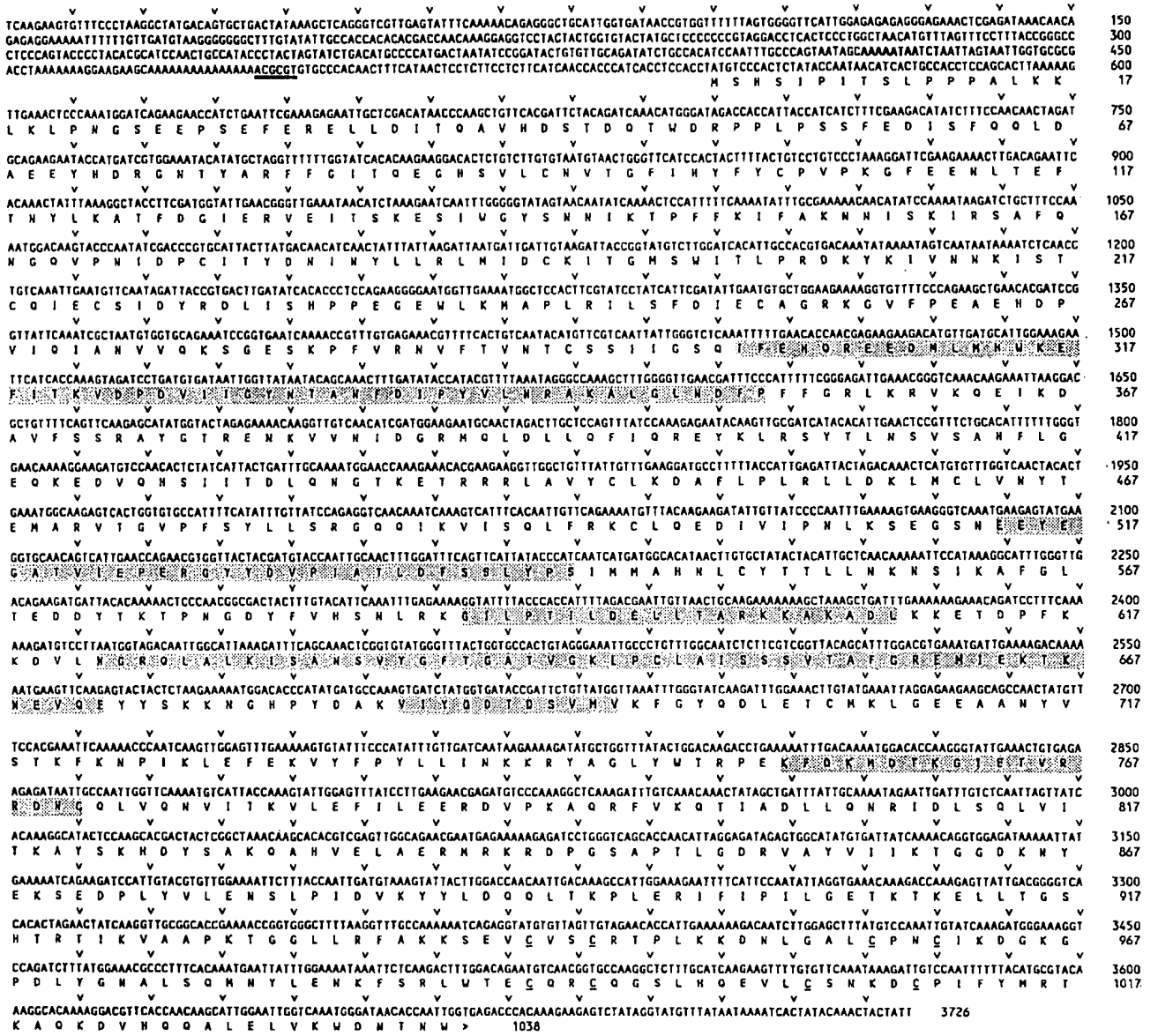

Fig. 2. Nucleotide sequence and predicted amino acid sequence of *C. albicans POL3* gene and its protein product. The features indicated here include: evolutionarily conserved motifs characteristic of DNA polymerases indicated by shading; cysteine residues which may form a potential zinc finger motif between amino acid residues 942 and 1010 are underlined; a potential TATAAA-like box in the upstream non-translated sequence located 120 bp upstream of the initiation codon is indicated in bold; a putative MCB 63 bp upstream from the initiation codon is indicated by double underline. Nucleotide sequence was determined on both strands using Sequenase 2.0 for dideoxynucleotide chain terminator sequencing of double stranded plasmid DNA (Sanger et al., 1977; Hsiao, 1991). The sequence shown here is available from the EMBL, Genbank and DDBJ Databases under accession No. X88804.

a heterodimeric protein that regulates the expression of about 30 genes at the G1-S phase boundary in *S. cerevisiae* (reviewed in McIntosh, 1993). The *CDC2* promoter contains a single copy of the extended consensus binding site for this transcription factor located 165 bp upstream of the transla-

Fig. 3. (**A**) Amino acid sequence homology within the six conserved regions of DNA polymerases identified by Wong et al. (1988). The sequences shown are *C. albicans* Pol3 (CaPol3), *S. cerevisiae* Cdc2 (Boulet et al., 1989), *S. pombe* pol δ (Park et al., 1993), EBV DNA polymerase (Baer et al., 1984) and human DNA polymerase α (Wong et al., 1988). The numbering shown is for the *C. albicans* Pol3 protein. (**B**) Comparison of the N- and C-terminal regions of *C. albicans* Pol3 with the corresponding regions of *S. cerevisiae* Cdc2 (Boulet et al., 1989) and HSV DNA polymerase (Kouzarides et al., 1987), indicating the conserved motifs of the viral sub-family of DNA polymerases (shaded).

A

Region I

```
            686        696
CaPol3       VIYGDTD SVMV
Cdc2         VVYGDTD SVMV
polδ+        VIYGDTD SVMV
EBV          VIYGDTD SLFI
polα human   VIYGDTD SIMI
```

Region II

```
            514                                    544
CaPol3       EEYEGATVIE  PERGYYDVPI ATLDLSSLYP S
cdc2         DQYEGATVIE  PIRGYYDVPI ATLDFNSLYP S
polδ+        EQYEGATVIE  PIKGYYDTPI ATLDFSSLYP S
EBV          DGYQGATVIQ  PLSGFYNSPV LVVDFASLYP S
polα human   GAYAGGLVLD  PKVGFYDKFI LLLDFNSLYP S
```

Region III

```
            622                                                          672
CaPol3       N GRQLALKISA NSVYGFTGAT VGKLPCLAIS SSVTAFGREM IEKTKNEVQE
Cdc2         N GRQLALKISA NSVYGFTGAT VGKLPCLAIS SSVTAYGRTM ILKTKTAVQE
polδ+        N GRQLALKVSA NSVYGFTGAT NGRLPCLAIS SSVTSYGRQM IEKTKDVVEK
EBV          D KQQLAIKCTC NAVYGFTGVA NGLFPCLSIA ETVTLQGRTM LERAKAFVEA
polα human   D IRQKALKLTA NSMYGCLGFS YSRFYAKPLA ALVTYKGREI LMHTKEMVQ
```

Region IV

```
            302                                                          348
CaPol3       IF EHQREEDMLM HWKEFITKVD PDVIIGYNTA NFDIPYVLNR AKALGLNDFP
Cdc2         IF SHATEEEMLS NWRNFIIKVD PDVIIGYNTT NFDIPYLLNR AKALKVNDFP
polδ+        VY EFQNQAEMLS SWSKFVRDVD PDVLIGYNIC NFDIPYLLDR AKSLRIHNFP
EBV          VY EFPSELDMLY AFFQLIRDLS VEIVTGYNVA NFDWPYILDR ARHIYSINPA
polα human   VE VAATERTLLG FFLAKVHKID PDIIVGHNIY GFELEVLLQR INVCKAPHWS
```

Region V

```
            832                    849
CaPol3       KFD...KMDTKG IETVRRDNC
Cdc2         KFD...KLDQKG LASVRRDSC
polδ+        TYD...KMDSKG IETVRRDNC
EBV          TDG...KTLMKG VELVRKTAC
polα human   DGNYVTKQELKG LDIVRRDWC
```

Region VI

```
            589                     609
CaPol3       GILP TILDELLTAR KKAKADL
Cdc2         GILP IILDELISAR KRAKKDL
polδ+        GLLP IILADLLNAR KKAKADL
EBV          SFLA SLLTSWLAKR KAIKKLL
polα human   GILP REIRKLVERR KQVKQLM
```

B

```
HSV     176 TPTGTVITLLGLTPEGHR.VAVHVYGTRQYFYMNKEEVDRHLQCRAP
Cdc2    131 ENTSTVVRFFGVTSEGH.SVLCNVTGFKNYLYVPAPNSSDANDQEQI
CaPol3   73 DRGNTYARFFGITQEGH.SVLCNVTGFIHYFYCPVPKGFEENLTEFT

HSV     350 LAIEGGMSDLPAYKLMCFDIECKAGGEDELAFPVAGHPEDLVIQISCLLYDLS
Cdc2    303 HPAEGDWSHTAPLRIMSFDIECAG...RIGVEP..EPEYDPVIQIANVVYLAL
CaPol3  232 HPPEGEWLKMAPLRILSPDIECAG...RKGVEP..EAEHDPVIQIANVVQKSG

HSV    1044 YTNKRLAHLTVYYKLMARR.AQVPSIKDRIPYVIVAQ
Cdc2    894 YTNPQ.PHAVLAERMKRRE.GVGPNVGDRVDYYIIGG
CaPol3  826 YSAKQ.AHVELAERMRKRDPGSAPTLGDRVAYVIIKT

HSV    1141 ELAEDPAYAIAHGVALNTDYYFSHLLGAACV
Cdc2    934 NRAEDPLFVLENNIQVDSRYYLTNQLQNPII
CaPol3  868 EKSEDPLYVLENSLPIDVKYYLDQQLTKPLE
```

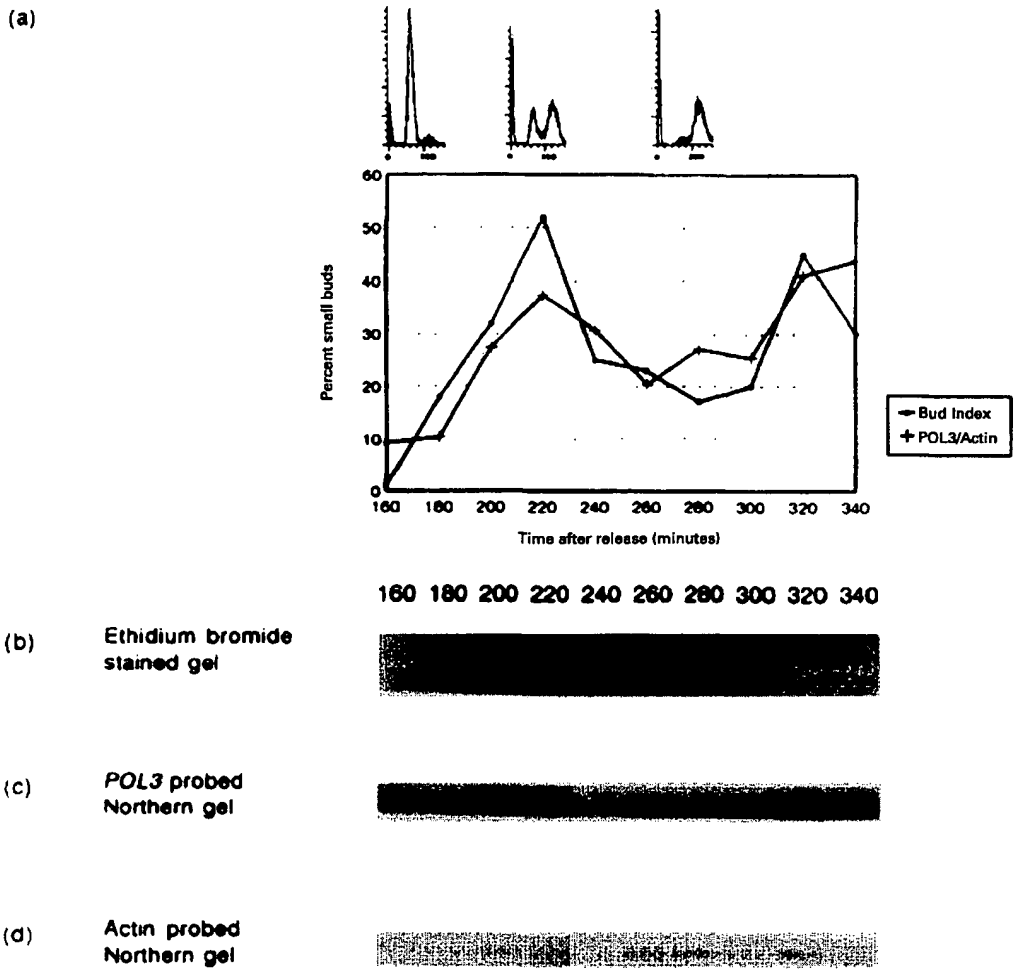

Fig. 4. Expression of *C. albicans POL3* through the mitotic cell cycle. *C. albicans* cultures were arrested by incubating overnight at 23°C in Edinburgh minimal medium without ammonium chloride (Nurse, 1975). Cells were harvested, resuspended in complete Edinburgh medium (pH 5.5) and grown with vigorous aeration at 23°C. (a) Aliquots were taken at 20 min intervals after release and analysed by FACS. Samples taken at 0 min (upper left), 220 min (upper centre) and 280 min (upper right) after release are shown. Samples at various times were also examined by microscopy for the appearance of small buds, which were scored as spherical outgrowths with a diameter less than a quarter that of the mother cell. This was plotted against the ratio of *POL3*/actin transcripts expressed in arbitrary units. Total RNA was prepared from samples of cells (Schmitt et al., 1990), resolved through agarose (Sambrook et al., 1989) and (b) visualised by staining with ethidium bromide; (c) probed for *C. albicans POL3;* and (d) probed for *S. cerevisiae* actin using the protocol of Church and Gilbert (1984). Filters were washed at 65°C and bound probe detected by autoradiography and phosphorimaging. For *C. albicans POL3*, the probe corresponded to a 397-bp region of 3′ end of *POL3* (nt 3318–3715) and was prepared by PCR.

tional start site (Boulet et al., 1989). A similar sequence is present in the promoter of *POL3*, 63 bp upstream of the initiation codon. However, we have found no evidence for the periodic accumulation of the *POL3* transcript. This may reflect the fact that the single MBF binding site acts principally to influence the level of synthesis rather than to mediate periodic expression (Lowndes et al., 1991; Verma et al., 1992). Alternatively, it may reflect the fact that *C. albicans* lacks transcription factors like MBF and so has no intrinsic mechanism for the periodic regulation of transcription at the G1-S phase transition. More likely though is the possibility that *C. albicans* contains proteins that resemble the components of MBF and which function in an analogous fashion, but that the number of genes regulated by this factor is considerably less than found in *S. cerevisiae*. In this regard, it is possible that *C. albicans* may more closely resemble *S. pombe*, where relatively few genes have been shown to be subject to regulation by the analogous transcription factor (McIntosh, 1993). Further analysis of genes encoding proteins required for DNA synthesis in *C. albicans* should clarify this point.

Acknowledgement

We would like to thank Anne White, Gavin Sherlock, Penny Johnson, Jeannette Hartley and Tracy Millar for practical assistance, suggestions and critical reading of the manuscript. T.N. is supported by a studentship from BBSRC.

References

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.J., Farrell, P., Gibson, T.J., Hatfull,G., Hudson, G.S., Satchwell, S.C., Sequin, C., Tuffnell, P. and Barrell, B.G. (1984) DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310, 207–211.

Bauer, G.A. and Burgers, P.M.J. (1990) Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucleic Acids Res. 18, 261–265.

Boulet, A., Simon, M., Faye, G., Bauer, G.A. and Burgers, P.M.J. (1989) Structure and function of the *Saccharomyces cerevisiae CDC2* gene encoding the large subunit of DNA polymerase III. EMBO J. 8, 1849–1854.

Church, G.M. and Gilbert, W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA 81, 1991–1995.

Culotti, J. and Hartwell, L.H. (1971) Genetic control of the cell division cycle in yeast. 3. Seven genes controlling nuclear division. Exp. Cell Res. 67, 389–401.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988) High efficiency transformation of *E.coli* by high voltage electroporation. Nucleic Acids Res. 16, 6127–6145.

Forsburg, S.L. and Nurse, P. (1991) Cell cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Ann. Rev. Cell Biol. 7, 227–256.

Gietz, R.D., St. Jean, A., Woods, R.A. and Schiestl, R.A. (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. 20, 1425.

Hsiao, K. (1991) A fast and simple procedure for sequencing double stranded DNA with Sequenase. Nucleic Acids Res. 19, 2787.

Jakab, E., Schmidt, T. and Hernadi F. (1991) Purification and some properties of *Candida albicans* DNA polymerases. Prep. Biochem. 21, 105–123.

Koch, C., Moll, T., Neuberg, M., Ahorn, H. and Nasmyth, K. (1993) A role for the transcription factors Mbp1 and SW14 in progression from G1 to S phase. Science 261, 1551–1557.

Kouzarides, T., Bankier, A.T., Satchwell, S.C., Weston, K., Tomlinson, P. and Barrell, B.G. (1987) Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene. J. Virol. 61, 125–133.

Lowndes, N.F., Johnson, A.L. and Johnston, L.H. (1991) Coordination of expression of DNA synthesis genes in budding yeast by a cell cycle regulated *trans* factor. Nature 350, 247–250.

McIntosh, E.M. (1993) MCB elements and the regulation of DNA replication genes in yeast. Curr. Genet. 24, 185–192.

Mitchell, L.H. and Soll, D.R. (1979) Commitment to germ tube or bud formation during release from stationary phase in *Candida albicans*. Exp. Cell Res. 120, 167–179.

Nasmyth, K. (1993) Control of the yeast cell cycle by the Cdc28 protein kinase. Curr. Opin. Cell Biol. 7, 227–256.

Nurse, P. (1975) Genetic control of cell size at cell division in yeast. Nature 256, 547–551.

Odds, F.C. (1988) *Candida* and Candidosis: A Review and Bibliography. Bailliere Tindall, London.

Park, H., Francesconi, S. and Wang, T.S. (1993) Cell cycle expression of two replicative DNA polymerase alpha and delta from *Schizosaccharomyces pombe*. Mol. Biol. Cell. 4, 145–157.

Patterson, M., Sclafani, R.A., Fangman, W.L. and Rosamond, J. (1986) Molecular characterisation of cell cycle gene *CDC7* from *Saccharomyces cerevisiae*. Mol. Cell. Biol. 6, 1590–1598.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Santos, M.A.S. and Tuite, M.F. (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res. 23, 1481–1486.

Schiestl, R.A. and Gietz, R.D. (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr. Genet. 16, 339–346.

Schmitt, M.E., Brown, T.A. and Trumpower, B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res. 18, 3091.

Schwob, E., Bohm, T., Mendenhall, M.D. and Nasmyth, K. (1994) The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae. Cell 79, 233–244.

Sherlock, G. and Rosamond, J. (1993) Starting to cycle: G1 controls regulating cell division in budding yeast. J. Gen. Microbiol. 139, 2531–2541.

Sherlock, G., Bahman, A.M., Mahal, A., Shieh, J-C., Ferreira, M. and Rosamond, J. (1994) Molecular cloning and analysis of *CDC28* and cyclin homologues from the luman fungal pathogen *Candida albicans*. Mol. Gen. Genet. 245, 716–723.

Sherman, F., Fink, G.R. and Hicks, J. (1986) Methods in Yeast Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122, 19–27.

Tyers, M., Tokiwa, G. and Futcher, B. (1993) Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. EMBO J. 12, 1955–1968.

Verma, R., Smiley, J., Andrews, B. and Campbell, J.L. (1992) Regulation of the yeast DNA replication genes through the *MluI* cell cycle box is dependant on *SWI6*. Proc. Natl. Acad. Sci. USA 89, 9479–9483.

Wong, S.W., Wahl, A.F., Yuan, P.-M., Arai, N., Pearson, B.E., Arai, K.-I., Korn, K., Hunkapiller, M.W. and Wang, T.S.-F. (1988) Human DNA polymerase a gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J. 7, 37–47.

Zhang, S.-J., Zeng, X.-R., Zhang, P., Toomey, N.L., Chuang, R.-Y., Chang, L.-S. and Lee, M.Y.W.T. (1995) A conserved region in the amino terminus of DNA polymerase δ is involved in proliferating cell nuclear antigen binding. J. Biol. Chem. 270, 7988–7992.

ELSEVIER

Gene 236 (1999) 43–51

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Cloning, sequencing, expression and allelic sequence diversity of *ERG3* (C-5 sterol desaturase gene) in *Candida albicans*

Yoshitsugu Miyazaki [a,1], Antonia Geber [b], Haruko Miyazaki [1,a], Derek Falconer [c], Tanya Parkinson [c], Christopher Hitchcock [c], Brian Grimberg [a], Katherine Nyswaner [a], John E. Bennett [a,*]

[a] *Clinical Mycology Section, Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, NIH. 10 Center Drive, Bethesda, MD 20892, USA*
[b] *Center for Biological Evaluation and Research, FDA, Rockville, MD, USA*
[c] *Department of Discovery Biology, Pfizer Central Research, Sandwich, UK*

Received 18 March 1999; received in revised form 19 May 1999; accepted 15 June 1999; Received by V. Larionov

Abstract

The C-5 sterol desaturase gene (*ERG3*), essential for yeast ergosterol biosynthesis, was cloned and sequenced from *Candida albicans* by homology with the *Saccharomyces cerevisiae ERG3*. The *ERG3* ORF contained 1158 bp and encoded 386 deduced amino acids. The clone was used to transform a *gal1* mutant derived from the Darlington strain of *C. albicans*, using galactose selection. The Darlington strain is known to lack $\Delta^{5,6}$ sterols, i.e. to have an *erg3* phenotype (Howell, S.A., et al., 1990. J. Appl. Bacteriol. 69, 692–696). The transformant (CDTR1) contained six tandem integrated *ERG3GAL1* repeats, had double the abundance of *ERG3* transcript found in the host strain, and synthesized ergosterol, a $\Delta^{5,6}$ sterol.

The Darlington strain was noted to have an abundance of *ERG3* transcript. Both *ERG3* alleles in Darlington were cloned and sequenced in order to look for changes that might explain the *erg3* phenotype. One allele, called Dar-2, contained a stop codon in place of tryptophan-292. The other *ERG3* allele, called Dar-1, had changes in three amino acids, two of which were conserved in three fungal and one plant species. *EcoRI* genomic fragments containing *ERG3* from the Dar-1 allele and from B311, the wild-type strain, were inserted into the plasmid pRS316 and used to transform a *Saccharomyces cerevisiae erg3,ura3* mutant using uracil selection. The 4.1 kb *ERG3* fragments from the B311 and Dar-1 both contained 1.4 kb 5′ and 1.5 kb 3′ flanking sequences around the coding region. Transformants with *ERG3* from B311 but not from Dar-1 showed restored ergosterol synthesis. One or more of these three deduced amino acids in the Dar-1 allele of *ERG3* appeared critical for function.

Keywords: Darlington; Ergosterol; *GAL1*; Heterozygosity; Yeast

1. Introduction

The enzyme, C-5 sterol desaturase, acts on C-5,6 saturated sterols such as episterol and ergosta-7, 22-dien-3β-ol at or near the end of the synthetic pathway for ergosterol (ergosta-5,7,22-trien-3β-ol) (Daum, et al. 1998). Mutants of *ERG3*, which encodes the C-5 sterol desaturase, have C-5,6 saturated bonds and, while retaining aerobic viability, have reduced growth rate (Taylor et al., 1983; Geber et al.,1995). Cloning and sequencing of *ERG3* has only been reported in *Saccharomyces cerevisiae* (Arthington et al, 1991; Hemmi et al., 1995), *Arabidopsis thaliana* (Gachotte et al., 1996) and by us in *Candida glabrata* (Geber et al., 1995). We extended our study to the *Candida albicans ERG3* because of the species' pathogenic potential and the possible relevance of *ERG3* to resistance against antifungal azoles. In *S. cerevisiae*, the *erg3* mutation has been postulated to suppress azole lethality by blocking accumulation of 14α-methyl-ergosta-8,24(28)-diene 3β,6α-diol, a putatively toxic sterol (Watson, et al. 1989). Having cloned the *C. albicans ERG3* by homol-

Abbreviations: 2DG, 2-deoxygalactose; bp, base pair; EDTA, ethylene diamine tetraacetic acid; kb, kilobase pair; Tris, tris(hydroxymethyl)aminomethane.

* Corresponding author. Tel.: +1-301-496-3461; fax: +1-301-480-0050.
E-mail address: jb46y@nih.gov (J.E. Bennett)
[1] Present address: Second Department of Internal Medicine, Nagasaki University School of Medicine, Nagasaki, Japan.

0378-1119/99/$ – see front matter
PII: S0378-1119(99)00263-2

ogy, we verified enzymatic function by complementing the only *erg*3 strain of *C. albicans* known to us. This strain, called Darlington, has not been mutagenized but obtained from a patient who received long-term azole therapy (Warnock et al., 1983). Darlington was known to produce only C-5,6 saturated sterols instead of ergosterol (Howell et al., 1990). No earlier isolate from this patient, without the *erg*3 mutation, is available. Because so little is known about the *ERG*3 gene, we studied the genetic basis of Darlington's *erg*3 phenotype.

2. Materials and methods

2.1. Strains and media

All strains used in this study are listed in Table 1. Yeast strains were maintained on YEPD medium [1% yeast extract, 2% peptone (DIFCO, Detroit, MI), and 2% glucose] unless noted. A *gal*1 mutant (DR16) of Darlington was selected by growth on 2-deoxygalactose (2-DG) plates [1% yeast extract, 2% peptone, 2% glycerol, 0.2% 2-DG (Sigma, St. Louis, MO)] (Gorman et al., 1992). Cells of DR16, which had been electroporated with DNA from plasmids pCADSG or pGal1-3K (Table 2), were screened for transformants on MINGal plates (0.64% yeast nitrogen base without amino acids, 2% galactose, 0.01% glucose). In selecting *GAL*1 transformants of *C. albicans* 1161, MINGal medium was supplemented with 0.002% arginine, 0.003% lysine, and 0.04% serine. *S. cerevisiae* strains transformed with the plasmids containing *ERG*3 (pRS-CAAR or pRS-DAAR) were selected and maintained in minimal media (0.64% yeast nitrogen base without amino acids, 2% glucose). The *E. coli* strains containing plasmids were grown and maintained in LBAmp [1% NaCl, 1% bactotryptone (DIFCO), 0.5% yeast extract (DIFCO), 100 μg/ml ampicillin (AMRESCO, Solon, OH)] broth or on plates (2% agar with LBAmp).

2.2. Nucleic acid preparation from yeast

DNA was isolated from yeast cells as described previously (Fujimura and Sakuma, 1993). Restriction digests were done according the manufacturer's directions (New England Biolabs, Beverly, MA). RNA was extracted with the FastRNA Kit (Bio 101, Vista, CA) from log phase cells grown in YEPD broth.

Table 1
C. albicans strains used in this study

Strain	Genotype	Origin or reference
C. albicans strains		
B311		Wild type strain
1161	*MPA*1/*MPA*1, *lys*1/*lys*1, *ura*3/*ura*3, *gal*1/*gal*1, *arg*4/*arg*4, *ser*57/*ser*57	Gift from Dr B. Wong, Yale Univ., New Haven, CT Goshorn et al, 1992; Wong et al., 1995
Darlington	*erg*3/*erg*3, *GAL*1/*GAL*1	C. Hitchcock, Sandwich, UK (Warnock et al., 1983)
DR16	*erg*3/*erg*3, *gal*1/*gal*1	Darlington *gal*1 mutant, selected on 2-DG in this study
CDTR1	*ERG*3/*ERG*3, *GAL*1/*gal*1	DR16 transformed with pCADSG in this study
CDTRGal	*erg*3/*erg*3, *GAL*1/*gal*1	DR16 transformed with pGal1-3K in this study
S. cerevisiae strains		
WA1α	*ERG*3,*leu*2,*his*7,*ade*5,*ura*3	Gift from Dr Martin Bard
WA1aL-316-23a	*erg*3, *LEU*2,*his*7,*ade*5,*ura*3	Gift from Dr Martin Bard
W-CAD	*ERG*3, *LEU*2,*his*7,*ade*5,*ura*3	WA1aL316-23a transformed with pRS-CADS
W-DAD	*erg*3, *LEU*2,*his*7,*ade*5,*ura*3	WA1aL316-23a transformed with pRS-DADS

Table 2
Plasmids used in this study

Plasmid	Description	Origin (reference)
pBSK	Bluescript II SK+ vector	Stratagene
pCAAR	5.5 kb *EcoR*I fragment containing *ERG*3 from B311 inserted into pBSK.	this study
pCADS	4.1 kb *Acc*I–*EcoR*I fragment containing *ERG*3 from B311 inserted into pBSK.	this study
pCADSG	2.7 kb pGal1-3 fragment inserted into pCAAR.	this study
pDAAR	5.5 kb Eco*R*I fragment containing *ERG*3 from Darlington inserted into pBSK.	this study
pDADS	4.1 kb *Acc*I–*EcoR*I fragment containing *ERG*3 from Darlington inserted into pBSK	this study
pGal1-3K	2.7 kb *EcoR*I fragment from pYSK208, containing *GAL*1 complementing activity	this study
pYSK208	11 kb DNA fragment with *GAL*1 in yEP13	gift from B. Magee
pRS316	yeast shuttle vector	ATCC (Sikorski and Hieter, 1989)
pRS-CAAR	5.5 kb *EcoR*I fragment containing *ERG*3 from B311 inserted into pRS316	this study
pRS-DAAR	5.5 kb *EcoR*I fragment containing *ERG*3 from Darlington inserted into pRS316	this study

2.3. Probes

A 1083 bp *S. cerevisiae ERG3* probe, used for cloning *ERG3* from *C. albicans* B311, was made by PCR from genomic DNA with oligonucleotides, 5′-ATGGATTTGGTCTTAGAA-3′ and 5′-CTTCTTGGTATTTGGGTC-3′, based on the published *S. cerevisiae ERG3* sequence (Arthington et al., 1991). Two *ERG3* probes were prepared by PCR using plasmid DNA from pCADS as template (Table 2) and *Taq* DNA polymerase (Boehringer Mannheim, Indianapolis, IN). A 0.86 kb *ERG3* probe was obtained with primers based on *C. albicans* sequences: 5′-GCCAGATCAAACATTTTCAGAG-3′ and 5′-AAAATAGTCAATGGTCCC-3′. For the Southern analysis in Fig. 3, a 0.65 kg *ERG3* probe was obtained with the following primers, also based on *C. albicans* sequences: 5′-AAATTGCTAGTTATCAAG-3′ and 5′-CATGAATCATGACAGTCC-3′. Genomic DNA from *C. albicans* B311 was used as template for preparing a 0.83 kb *ACT1* probe by PCR using the following primers: 5′-TATCGATAACGGTTCTGG-3′ and 5′-CATCACACTTCATGATGG-3′ (Losberger and Ernst, 1989). A probe containing the 5′ end of the *C. albicans GAL1* open reading frame (ORF) was prepared by excising the 1.5 kb *Acc*I–*Eco*R1 fragment from pGAL1-3K. The location of the sequences selected for the probes is given in Fig. 1. Radioactive probes were prepared by using RTS RadPrime DNA Labeling System (Life Technologies, Gaithersburg, MD) or Prime-It II (Stratagene, La Jolla, CA) according to the manufacturers' instructions. Sequencing was performed with a rhodamine terminator sequencing reaction running on an ABI Prism 377 (Perkin Elmer, Foster city, CA) or

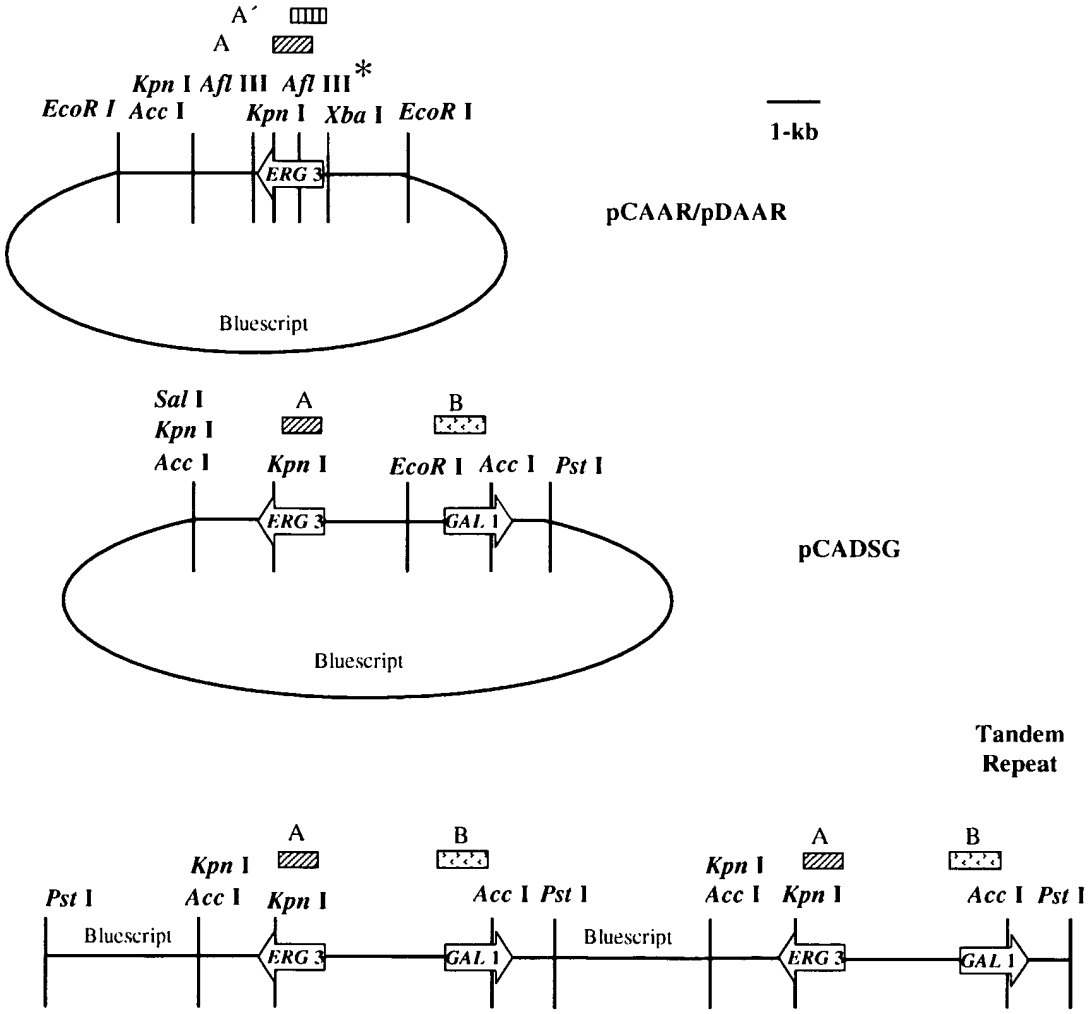

Fig. 1. Restriction maps for plasmids constructed in this study and for the tandem repeat integrated in CDTR1. The ellipse represents pBSK vector DNA, and open arrows designate ORFs of *ERG3* and *GAL1* with direction of transcription. Boxes show the sequences hybridized by the probes in Southern and northern analysis. 'A' denotes the 0.86 kb *ERG3* probe, 'A" the 0.65 kb *ERG3* probe and 'B' the *GAL1* probe. An asterisk designates the *Afl*III site that is present in Dar-1 and not Dar-2 or B311.

with dideoxy-DNA sequencing using the Sequenase version 2 sequencing kit (Amersham, Arlington Heights, IL). The nucleic acid sequences were analyzed by programs from the Genetics Computer Group, Madison, WI (Devereux et al., 1984). Deduced amino acid sequences were compared by Bestfit analysis, permitting up to four gaps (Devereux et al., 1984).

2.4. Southern and northern analyses

Northern and Southern analyses were performed as described (Sambrook et al., 1989). Probes were labeled with [^{32}P]-dCTP by the Prime-It II kit (Stratagene). Southern and colony blots were probed with the *S. cerevisiae ERG3*, hybridized at 42°C and finally washed at 48°C in washing solution-1 [2 × SSC (0.3 M NaCl plus 0.03 M sodium citrate), 0.1% SDS]. Southern blots probed with *C. albicans ERG3* or the *GAL1* probe were hybridized at 65°C and finally washed at 65°C in solution-2 (0.2 × SSC, 0.1% SDS).

2.5. Subcloning of the C. albicans GAL1

pYSK208 was digested with *EcoR*I (New England Biolabs), electrophoresed and the 2.7 kb fragment subcloned into the *EcoR*I site of pBSK, yielding pGAL1-3K (see Table 2). All plasmids were electroporated with the DH10B strain of *E. coli* (Life Technologies). Integrative transformation of *C. albicans* 1161, *gal*1 with pGAL1-3K restored galactose utilization.

2.6. Transformation of DR16

DR16 was transformed first with pGAL1-3K linearized with *EcoR*I. The transformant CDTRGal was selected for further study. DR16 was also transformed with pCADSG linearized with *Pst*I. The entire ligation mixtures were plated and incubated 5 days at 30°C. The number of transformants per milligram of DNA was >300 for pGAL1-3K and 61 for pCADSG. The no-DNA control yielded 25 colonies of *GAL1* backrevertants. Southern analysis of an *Acc*I digest, probed for *ERG3*, was used to select a pCADSG transformant, CDTR1 (data not shown).

2.7. Yeast transformation

Electroporation was done in *gal*1 *C. albicans* strains as published (Varma et al., 1992). Electroporated cells were spread on MINGal plates and incubated at 37°C for 3 to 7 days.

Lithium acetate (Gietz and Woods, 1994) was used to transform WA1aL-316-23a, an *erg3 S. cerevisiae* strain, with pRS-CAAR and pRS-DAAR (Table 2) using uracil selection. Confirmation that WA1aL-316-23a contained non-integrated copies of the *C. albicans ERG3* was obtained by DNA extraction, gel electrophoresis of undigested DNA, blotting and hybridization with the *C. albicans ERG3* probe.

2.8. CHEF

Pulse field electrophoresis was performed using 0.8% chromosomal grade agarose (BioRad, Hercules, CA) in 0.5 × TBE (Tris 45 mM, borate 45 mM, EDTA 1 mM, buffer pH 8.0) on 14 × 12.5 cm^2 gels using a BioRad electrophoresis chamber with a CHEF-DR II Drive Module (BioRad). Runs were done at 12°C with 150 V and 120 s switch times ramped to 240 s over 25 h followed by 180 s switch times ramped to 360 s over 20 h. The chromosomal DNA run in CHEF gels was first depurinated with 0.25 M HCl at room temperature for 20 min, then DNA was transferred to nylon membranes (Hybond-N, Amersham) for probing with *GAL1* as described in Section 2.3.

2.9. Quantification of transcription and copy number

Filters hybridized with radiolabeled probes were exposed to Storage Phosphor Screen (Molecular Dynamics, Sunnyvale, CA) for 3 h, and the screen was scanned on a PhosphoImager 445 SI (Molecular Dynamics). The images obtained were analyzed with quantification software, ImageQuant (Molecular Dynamics). Quantitative volume data for the same rectangular square on the blot image with each probe were used for analysis. Relative transcription levels were calculated from data obtained by the same blot with different probes.

2.10. Sterol analysis

Sterol identification by gas chromatography (GC) and GC–mass spectrometry of trimethylsilyl derivatives was done as published previously (Geber et al., 1995).

3. Results

3.1. Cloning and sequencing of ERG3 from C. albicans strain B311

The 1083 kb *S. cerevisiae ERG3* PCR product described in Section 2.3 was used to probe a Southern blot of *C. albicans* B311 DNA digested with restriction endonucleases. On the basis of these results, an *EcoR*I-digested *C. albicans* genomic library containing fragments in the 5 to 6 kb range was constructed in pBSK. Following transformation into *E. coli*, clones were screened with the *S. cerevisiae* probe and further identified by sequence analysis. pCAAR, obtained from this

library, contained a 5.5 kb insert and was further subcloned in pBSK to yield a 4.1 kb *EcoRI–AccI* fragment, pCADS (Fig. 1). Sequencing within pCADS found an 1158 bp ORF, encoding 386 deduced amino acids. A total of 2023 bases were sequenced, including 386 bp 5′ to ORF and 479 bases in the 3′ flanking region (Genbank accession no. AF069752).

There was no CUG codon, which codes for serine instead of leucine in *Candida* (Jute and Osawa, 1996). Two putative TATA boxes appeared 302 bp and 52 bp upstream of the ATG start codon. A ten-base sequence, TCGTTTAAGT was found 370 to 379 bp upstream to the initiation codon. This sequence differs by one base from the TCGTATAAGT at position 277 to 286 upstream from the *S. cerevisiae ERG3* coding region, part of the upstream activating sequence (UAS2) described by Arthington-Skaggs et al. (1996). However, no sequence homology was found with UAS1 (Arthington-Skaggs et al., 1996) or with the *ERG3* regulatory sequence at 390 to 412 bp upstream of the *S. cerevisiae ERG3* (Smith, et al., 1996). A transcription termination signal (Zaret and Sherman, 1982) was observed 307 bp downstream from the B311 *ERG3* TGA stop codon. On Bestfit analysis of the deduced *C. albicans* B311 peptide sequences, the *S. cerevisiae* and *C. glabrata* C-5,6 desaturases had 59.5% and 59.6% identity over 317 and 294 amino acids respectively. *A. thaliana* C-5,6-desaturase had only 34.1% identity over 186 amino acids. Hybridization of CHEF blots showed that both the *GAL1* and *ERG3* probes hybridized to the largest chromosome in B311 and Darlington (data not shown). Four putative iron binding domains were found (Fig. 2).

3.2. *Cloning and sequencing of* ERG3 *from* C. albicans *Darlington strain*

The same strategy used to construct pCAAR and pCADS from B311 genomic DNA was used to construct pDAAR and pDADS from genomic DNA of the Darlington strain. The *ERG3* ORF in pDADS was sequenced in its entirety with the same primers used for pCADS. The sequence revealed an *AFLIII* site in pDADS not present in pCADS. Southern blots were prepared from genomic DNA from Darlington and B311, doubly digested with *Xba*I and *Afl*III, and probed with the 0.65 kb *C. albicans ERG3* probe described in Section 2.3. B311 yielded only a 1.5 kb band, whereas Darlington had 1.5, 0.9 and 0.6 kb bands, consistent with an *Afl*III site found in only one Darlington allele, designated Dar-1, and not in B311 (Fig. 3). In order to clone the allele of Darlington that was not restricted by *Afl*III, designated Dar-2, a 1.4 kb PCR product containing the entire 1.1 kb *ERG3* coding sequence was obtained using Darlington genomic DNA as template and the following primers: 5′-AAAATAGTCAATGGTCCC-3′

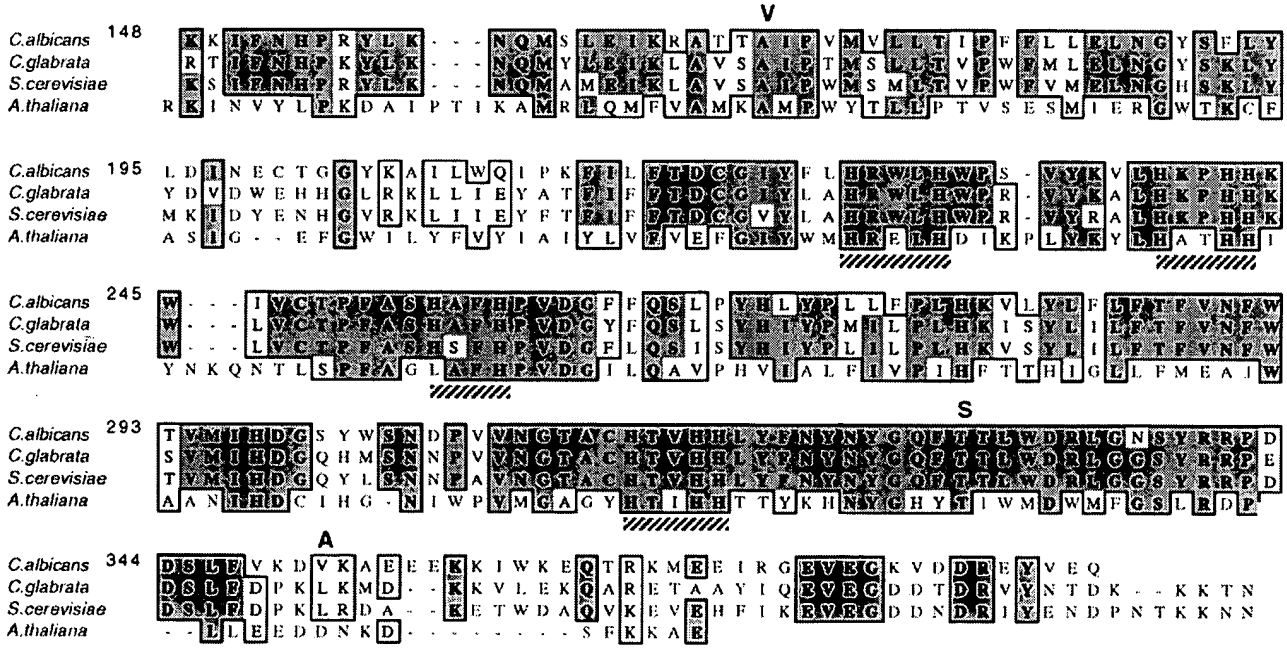

Fig. 2. Deduced amino acid sequences of C-5,6 desaturase in four different species were aligned. The first three are yeast species, *C. glabrata* (GenBank accession no. L40390), *S. cerevisiae* (GenBank accession no. M64989) and *C. albicans* (GenBank accession no. AF069752), and the fourth row is a plant, *A. thaliana* (EMBL accession no. A520G15F). Numbering was based on *C. albicans* sequence. Identical (emboldened) or similar amino acids in at least two species were boxed. Four histidine-containing putative metal-binding domains were underlined. Amino acids that differed in Dar-1 are designated above the arrays.

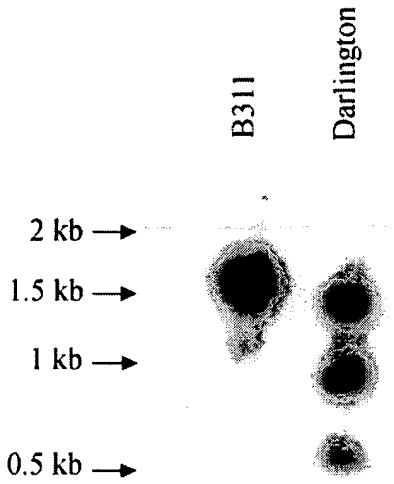

Fig. 3. Southern blot of B311 and Darlington genomic DNA double digest with *Xba*I and *Afl*III, probed with the 0.65 kb *ERG3* fragment.

and 5′-ACATTACTGCTTACTTTGAGAG-3′. When the 1.4 kb band was excised from the gel, extracted by GeneClean and restricted with *Afl*III, the 1.4 kb band persisted, in addition to the expected 0.7 kb fragments. The undigested 1.4 kb band was excised from the gel, the DNA extracted, and ligated into the *Srf*I site of PCR-Script (Stratagene). Several plasmids recovered from transformants had inserts that were not restricted by *Afl*III. Two more PCR reactions using Darlington genomic DNA as template revealed products that on insertion into PCR-Script also had fragments that were not restricted by *Afl*III. Nucleotide sequences of the 1.4 kb fragments in two of the plasmids from different reactions were identical, both having a stop codon in place of tryptophan-292. The difference in base and deduced amino acid sequence for *ERG3* in B311 and the two Darlington alleles is given in Table 3. In the Dar-1 sequence, three deduced amino acids were different from the wild-type strain, B311. Two of these amino acids were conserved in *S. cerevisiae* (Arthington et al., 1991), *C. glabrata* (Geber et al., 1995) and *A. thaliana* (Gachotte et al., 1996).

3.3. *Transformation of Darlington with B311* ERG3

Electroporation of DR16, a *gal*1 auxotroph of *C. albicans* Darlington, with 2 μg *Pst*I-linearized pCADSG yielded the transformant CDTR1, which was passed through MINGal plates several times to assure stability. Southern analysis of *Acc*I, *Kpn*I and *Acc*I–*Kpn*I digests (Fig. 4) showed the native *GAL*1 and *ERG3* genes to be of the predicted sizes, 3 kb and 5 kb respectively. The Southern analysis of CDTR1 indicated that pCADSG, including pBSK sequences, had integrated as tandem repeats at one or more sites different from *ERG3* and *GAL*1, consistent with the restriction map shown (Fig. 1). The increased intensity of the integrated *GAL*1–*ERG3* fragment was quantified by the PhosphoImager, using the native *ERG3* for comparison. The intensity of the 4.8 kb band was 3.15 times that of the native 3 kb band. Assuming that the native *ERG3* was diploid, the ectopically integrated *ERG3* existed as six copies.

3.4. ERG3 *transcription in* C. albicans *strains*

In northern analysis, DR16 had an *ERG3* transcript of expected size, approximately 2 kb. On PhosphoImager analysis of northern blots, CDTR1 showed an average of 2.1 times as much *ERG3* transcript as Darlington and 7.9 times as much as B311 using the transcript of the constitutively expressed *ACT*1 (Losberger and Ernst, 1989) as a control. Darlington had an average of 3.8 times as much *ERG3* transcript as B311 (Fig. 5.)

3.5. ERG3 *expression in* C. albicans *strains*

Sterols extracted from CDTR1 contained 95.3% ergosterol and no detectable ergosta-7,22-dien-3β-ol

Table 3
Sequence divergence of *ERG3* alleles in *C. albicans*

Base position[a]	*ERG3* alleles in each strain			Deduced amino acids			
	B311	Dar-1	Dar-2	#	B311	Dar-1	Dar-2
49	TAC	TA**T**	TAC	17	Tyr	Tyr	Tyr
304	ACT	ACT	AC**C**	102	Thr	Thr	Thr
430	TTC	TT**T**	TT**T**	144	Phe	Phe	Phe
502	GCC	G**T**C	GCC	168	Ala	Val	Ala
874	TGG	TGG	T**A**G	292	Trp	Trp	stop
985	ACT	A**G**T	ACT	319	Thr	Ser	Thr
1051	GTA	G**C**A	G**C**A	351	Val	Ala	Ala

[a] Position represents the first base of B311's ORF. Underlining shows the bases that differ between the Darlington alleles and B311.

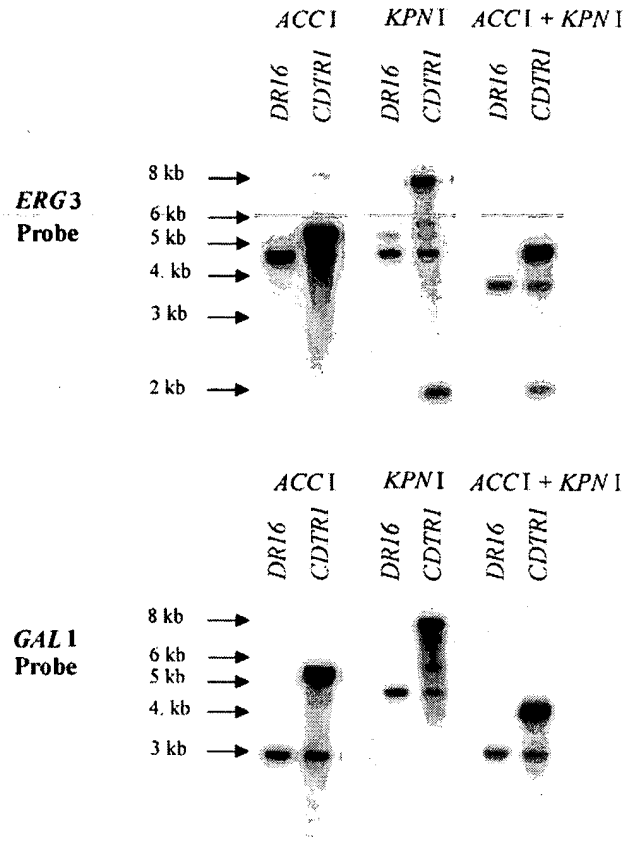

Fig. 4. Southern analysis of *C. albicans*. Southern analysis is shown of DR16 and CDTR1 DNA digested with the enzymes indicated and probed with *ERG3* and *GAL1* fragments. *ERG3* probing revealed a 4.8 kb native *Acc*I fragment in both strains and an additional 5.6 kb fragment only in the transformant, CDTR1. Blots of *Kpn*I digests probed with *ERG3* showed 8.6 and 1.6 kb fragments in addition to the native gene in CDTR1, as predicted from the map of pCADSG. Another band is seen that may represent a flanking region at which recombination occurred. On *GAL1* probing, an 8.6 kb fragment, representing the integrated sequence, was seen in the transformant. Two other bands are seen between the native gene and 8.6 kb fragments. One of the two was the same size as the fragment that hybridized with the *ERG3* probe and may represent flanking regions. The *Acc*I–*Kpn*I digest showed the predicted fragment of approximately 4 kb in CDTR1 using both probes.

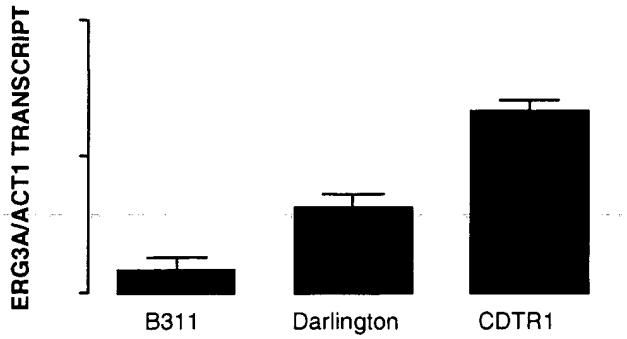

Fig. 5. Northern analysis of *C. albicans*. The quantity of *ERG3* transcript, as assessed by northern analysis, was compared between three strains: B311, Darlington and CDTR1. Data represent the mean ± SD of *ERG3*/*ACT1* transcript in three experiments.

(Table 4). The *C. albicans* Darlington *gal*1 host, DR16, contained 96.1% ergosta-7,22-dien-3β-ol and no detectable ergosterol. Sterol composition was similar when DR16 was transformed only with *GAL1* to yield CDTRGal. These results show that *ERG3* from B311 had complemented the *erg*3 mutation in Darlington.

3.6. ERG3 *expression in a* S. cerevisiae erg3 *mutant*

To elucidate if Darlington's *ERG3* homolog with three amino acid differences was non-functional, Dar-1 was compared with the B311 *ERG3* for ability to restore ergosterol synthesis in WA1aL-316-23a, a *S. cerevisiae erg*3 mutant. Four transformants from each electroporation were confirmed to carry the *C. albicans ERG3* gene as described in Section 2.7 and were analyzed for sterol pattern. The four (designated W-CAD1-4), with the *ERG3* gene from B311 inserted in pRS316, restored the *S. cerevisiae* host's ability to synthesize ergosterol. Ergosterol constituted 36.9, 38.6, 59.4 and 64.5% of sterols extracted from these transformants (Table 4). The host strain WA1aL-316-23a, had no detectable ergosterol and the parental strain, WA1α, had 62.6%. None of the four transformants with the Dar-1 *ERG3* (designated W-DAD1-4) had detectable ergosterol.

4. Discussion

We cloned and sequenced the *ERG3* gene from *C. albicans* and found the gene to be located on the same chromosomal CHEF band as *GAL1*, which is chromosome I. The sequence contains a putative TATA box (TATAA) and the same translation terminator signal as *S. cerevisiae* (Zaret and Sherman, 1982). A *gal*1 mutant was selected on 2-DG medium and used as a selection marker to transform Darlington with *ERG3* from a stock *C. albicans* strain, B311, to see if ergosterol synthesis could be restored. We were able to show that Darlington was an *erg*3 mutant, as predicted (Howell et al., 1990), and that the *ERG3* clone complemented Darlington by changing the major sterol to ergosterol.

Sequencing the Darlington *ERG3* found allelic heterozygosity. One allele, Dar-2, had a stop codon in midsequence and the other allele, Dar-1, differed from strain B311 in three deduced amino acids. Failure of Dar-1 to complement a *S. cerevisiae erg*3 mutant, using a system in which *ERG3* from B311 was successful, showed the importance of at least one of these amino acids for its function. Since A168 and T329 were highly conserved, not only in yeasts but also in a plant species, *A. thaliana* (Fig. 2), these two amino acids were considered more likely to be critical for *ERG3* function. Transcriptional

Table 4
Sterol components in yeast strains[a]

	Zymosterol	Fecosterol	Ergosta-8-en-3β-ol	Ergosta-7-en-3β-ol	Ergosta-7,22-dien-3β-ol	Ergosterol	Unknown
C. albicans strains							
DR16	–	–	2.0	–	96.1	–	–
CDTR1	4.7	–	–	–	–	95.3	–
CDTRGal	–	1.2	6.3	–	90.7	–	–
S. cerevisiae strains							
Wa1-α	5.5	11.5	–	10.8	–	62.6	9.6
WA1aL316-23a	–	14.4	12.7	–	72.6	–	–
W-CAD1	7.6	4.1	14.8	5.3	26.1	36.9	5.1
W-CAD3	9.6	5.9	11.6	5.9	22.2	38.6	5.7
W-CAD4	3.6	3.9	6.4	5.4	12.2	64.5	4.0
W-CAD6	3.1	4.9	10.2	4.2	14.2	59.4	3.2
W-DAD5	–	16.2	10.0	6.8	66.8	–	–
W-DAD11	–	14.6	11.2	4.8	69.4	–	–
W-DAD12	–	15.3	13.4	5.2	66.0	–	–
W-DAD17	–	–[b]	–[b]	–	78.2	–	–

[a] Except where indicated, a bar (–) represents a sterol amount that was no more than 1%.
[b] This sample included approximately 22% of fecosterol and ergosta-8-en-3β-ol, which were not clearly separated.

regulation is an unlikely explanation for the phenotype, considering the abundance of a transcript of the correct size. However, the cDNA was not cloned. Nor was transcriptional regulation studied.

Allelic sequence divergence was not surprising in *C. albicans*. Higher levels of sequence divergence are expected to be seen in species with no known perfect state than in organisms that mate (Birky, 1996). Kelly et al. (1987) were the first to describe restriction fragment polymorphism in a *C. albicans* gene, which was an *Eco*RI site in *URA*3. Other examples of restriction site polymorphism in genetic loci in this species have been reported, such as the *Hind*III site of *ERG*11 (Kirsch et al., 1988). Allelic sequence diversity has been noted in *SAP*4 (Miyasaki et al., 1994) and *ERG*11 (White, 1997). The current report may be the first report in *C. albicans* of a functionally obvious difference between alleles, i.e. a stop codon in the middle of the ORF.

When Pujol et el. (1993) examined 21 gene loci of 55 *C. albicans* strains using multilocus enzyme electrophoresis, the mean heterozygosity was 0.168 with a range of 0.013 to 0.430. This technique shows that heterozygosity is quite common, but the method cannot distinguish whether the alleles are at the same or different loci. Auxotrophic heterozygosity in wild-type strains has been used to explain why certain auxotrophs have been much easier than others to obtain during mutagenesis in this species, although this provides indirect evidence (Kirsch, 1990). Chromosomal rearrangements are one mechanism potentially leading to heterozygosity based on ability of the chromosomal alteration to be associated with a change in morphotype (Rustchenko-Bulgac et al., 1990) or assimilation pattern (Rustchenko et al., 1994). The Darlington strain, although a wild type, was under selective pressure from prolonged azole antifungal therapy. The effect may have been analogous to the chromosomal aberrations that occurred when certain *C. albicans* strains were grown on sorbose (Pujol et al., 1993).

Acknowledgements

We thank Dr K.J. Kwon-Chung for helpful discussions and Dr Martin Bard and Dr Scott Moye-Rowley for their advice and gifts of fungal strains. This work was supported in part by a grant from Pfizer, Inc.

References

Arthington, B.A., Bennett, L.G., Skatrud, P.L., Guynn, C.J., Barbuch, R.J., Ulbright, C.E., Bard, M., 1991. Cloning, disruption and sequence of the gene encoding yeast C-5 sterol desaturase. Gene 102, 39–44.

Arthington-Skaggs, B.A., Crowell, D.N., Yang, H., Sturley, S.L., Bard, M., 1996. Positive and negative regulation of a sterol biosynthesis gene (*ERG*3) in the post-squalene portion of the yeast ergosterol pathway. FEBS Lett. 392, 161–165.

Birky Jr, C.W., 1996. Heterozygosity, heteromorphism and phylogenetic trees in an asexual eukaryote. Genetics 144, 427–437.

Daum, G., Lees, N.D., Martin, M., Dickson, R., 1998. Biochemistry, cell biology and molecular biology of lipids of *Saccharomyces cerevisiae*. Yeast 14, 1471–1510.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Fujimura, H., Sakuma, Y., 1993. Simplified isolation of chromosomal and plasmid DNA from yeast. BioTechniques 14, 538–539.

Gachotte, D., Hasselstein, T., Bard, M., Lacroute, F., Benveniste, P., 1996. Isolation and characterization of an *Arabidopsis thaliana* cDNA encoding a Δ7-sterol-C-5-desaturase by functional complementation of a defective yeast mutant. Plant J. 9, 391–398.

Geber, A., Hitchcock, C.A., Swartz, J.E., Pullen, F.S., Marsden, K.E., Kwon-Chung, K.J., Bennett, J.E., 1995. Deletion of the *Candida glabrata ERG*3 and *ERG*11 genes: effect on cell viability, cell

growth, sterol composition and antifungal susceptibility. Antimicrob. Agents Chemother. 39, 2708–2717.

Gietz, R.D., Woods, R.A., 1994. High efficiency transformation in yeast. In: Johnston, J.W. (Ed.), Molecular Genetics of Yeast: Practical Approaches. Oxford University Press, pp. 121–134.

Gorman, J.A., Gorman, J.W., Koltin, Y., 1992. Direct selection of galactokinase-negative mutants of *Candida albicans* using 2- deoxygalactose. Curr. Genet. 21, 203–206.

Goshorn, A.K., Grindle, S.M., Scherer, S., 1992. Gene isolation by complementation in *Candida albicans* and applications to physical and genetic mapping. Infect. Immun. 60, 876–884.

Hemmi, K.C., Julmanop, C., Hirata, D., Tsuchiya, E., Takemoto, J.Y., Miyakawa, T., 1995. The physiological roles of membrane ergosterol as revealed by the phenotypes of *syr1/erg3* null mutants of *Saccharomyces cerevisiae*. Biosci. Biotech. Biochem. 59, 482–486.

Howell, S.A., Mallet, A.I., Noble, W.C., 1990. A comparison of the sterol content of multiple isolates of the *Candida albicans* Darlington strain with other clinically azole-sensitive and -resistant strains. J. Appl. Bacteriol. 69, 692–696.

Jute, T.H., Osawa, S., 1996. CUG codons in *Candida* spp. J. Mol. Evolution 42, 321–322.

Kelly, R., Miller, S.M., Kurtz, M.B., Kirsch, D.R., 1987. Directed mutagenesis in *Candida albicans*: one-step gene disruption to isolate *ura3* mutants. Mol. Cell. Biol. 7, 199–208.

Kirsch, D.R., Kelly, R., Kurtz, M.B., 1990. The Genetics of *Candida*. CRC Press, Boca Raton, FL.

Kirsch, D.R., Lai, M.H., O'Sullivan, J., 1988. Isolation of the gene for cytochrome P450L1A1 (lanosterol 14alpha-demethylase) from *Candida albicans*. Gene 68, 229–237.

Losberger, C., Ernst, J.F., 1989. Sequence of the *Candida albicans* gene encoding actin. Nucleic Acids Res. 17, 9488

Miyasaki, S.H., White, T.C., Agabian, N., 1994. A fourth secreted aspartyl proteinase gene (*SAP4*) and a *CARE2* repetitive element are located upstream of the *SAP1* gene in *Candida albicans*. J. Bacteriol. 176, 1702–1710.

Pujol, C., Reynes, J., Renaud, F., Raymond, M., Tibayrenc, M., Ayala, F.J., Janbon, F., Mallie, M., Bastide, J.M., 1993. The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. Proc. Natl. Acad. Sci. USA 90, 9456–9459.

Rustchenko, E.P., Howard, D.H., Sherman, F., 1994. Chromosomal alterations of *Candida albicans* are associated with the gain and loss of assimilating functions. J. Bacteriol. 176, 3231–3241.

Rustchenko-Bulgac, E.P., Sherman, F., Hicks, J.B., 1990. Chromosomal rearrangements associated with morphological mutants provide a means for genetic variation of *Candida albicans*. J. Bacteriol. 172, 1276–1283.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual. 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sikorski, R.E., Hieter, P., 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122, 19–27.

Smith, S.J., Crowley, J.H., Parks, L.W., 1996. Transcriptional regulation by ergosterol in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. 16, 5427–5432.

Taylor, F.R., Rodriguez, R.J., Parks, L.W., 1983. Requirement for a second sterol biosynthetic mutation for viability of a sterol C-14 demethylation defect in *Saccharomyces cerevisiae*. J. Bacteriol. 155, 64–68.

Varma, A., Edman, J.C., Kwon-Chung, J.K., 1992. Molecular and genetic analysis of *URA5* transformants of *Cryptococcus neoformans*. Infect. Immun. 60, 1101–1108.

Warnock, D.W., Johnson, E.M., Richardson, M.D., Vickers, C.F.H., 1983. Modified response to ketoconazole of *Candida albicans* from a treatment failure. Lancet i, 642–643.

Watson, P.E., Rose, M.E., Ellis, S.W., England, H., Kelly, S.L., 1989. Defective sterol C5-6 desaturation and azole resistance: a new hypothesis for the mode of action of azole antifungals. Biochem. Biophys. Res. Commun. 164, 1160–1175.

Wong, B., Leeson, S., Grindle, S., Magee, B., Brooks, E., Magee, P.T., 1995. D-Arabitol metabolism in *Candida albicans*: construction and analysis of mutants lacking D-arabitol dehydrogenase. J. Bacteriol. 177, 2971–2976.

White, T.C., 1997. The presence of an R467K amino acid substitution and loss of allelic variation correlate with an azole-resistant lanosterol 14-α demethylase in *Candida albicans*. Antimicrob. Agents Chemother. 41, 1488–1494.

Zaret, K.S., Sherman, F., 1982. DNA sequence required for efficient transcription termination in yeast. Cell 28, 563–573.

Nucleic Acids Research, 1996, Vol. 24, No. 14 2627–2631

Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions

Jean-Pierre Vartanian, Michel Henry and Simon Wain-Hobson*

Unité de Rétrovirologie Moléculaire, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris cedex 15, France

Received May 7, 1996; Revised and Accepted June 5, 1996

ABSTRACT

Very complex mutant libraries of the dihydrofolate reductase (DHFR) gene encoded by the *Escherichia coli* plasmid R67 were created using hypermutagenic PCR with biased deoxynucleotide triphosphate (dNTP) concentrations. Exploiting the particular stability of the G:T mismatch, the DHFR gene could be enriched in A+T by employing biased deoxypyrimidine triphosphate concentrations, i.e. [dTTP] > [dCTP]. A sizeable fraction of hypermutants were functional. A combination of [dTTP] > [dCTP] and [dGTP] > [dATP] biases generated mutations at unexpectedly low frequencies. This could be overcome by the addition of Mn^{2+} cations. Overall mutation frequencies of 10% per amplification (range 4–18% per clone) could be attained. All four transitions and a smaller number of transversions were produced throughout the gene. PCR mutagenesis could be so extensive as to inactivate all amplified versions of the gene.

INTRODUCTION

Although the mutation rates of DNA based organisms vary, they are considerably less than one per genome per cycle. Those of the RNA viruses may approach two to four substitutions per genome per cycle (1). Such rates must represent the upper end of the spectrum compatible with viability as they may be only slightly increased by chemical mutagenesis (2). Higher mutation rates almost certainly result in extinction. However, apart from this obvious restriction there is nothing *per se* to prohibit higher mutation rates *in vitro* or hypermutation restricted to small regions of a genome or gene segment *in vivo* (3). Perhaps the most startling example of this is retroviral G→A hypermutation where hundreds of templated Gs may be copied into As (4–6). This is a particular trait of the lentiviral family of retroviruses, which includes human immunodeficiency virus (HIV), and results from cDNA synthesis in the presence of highly biased [dTTP]/[dCTP] ratios (6).

G→A hypermutation can be reproduced *in vitro* using RNA, biased dNTP concentrations and preferentially the HIV-1 reverse transcriptase (7–9). Referred to as RNA hypermutagenesis, this method delivers elevated mutation and mutant frequencies, ≤0.1 per G per cycle and >0.9 per DHFR gene per cycle respectively. The complexity of the resulting libraries of hypermutated sequences was limited by the monotony of G→A hypermutation. Despite this, iterative hypermutagenesis of a bacterial antibiotic resistance gene, the *Escherichia coli* R67 DHFR, resulted in substitution of up to 23% of amino acids without loss of phenotype (10).

Genes and genomes exhibit G+C- or A+T-rich segments so that it would be useful to have a method capable of enriching any sequence in either. Just as dNTP biases are mutagenic for reverse transcription (7,11) so they are for PCR (12–16), although the magnitude of the bias has to be less to allow reasonably efficient amplification. PCR has the advantage that both strands may be mutated. A [dTTP] > [dCTP] bias would allow enrichment in A and T while a [dGTP] > [dATP] bias would permit the converse. These biases generate $G_{t(template)}$:T and T_t:G mismatches respectively which are the most stable of the 12 possible (17). By combining both deoxypyrimidine and deoxypurine triphosphate biases, it is shown here that PCR can be hypermutagenic to an unprecedented degree.

MATERIALS AND METHODS

The oligonucleotides used for amplification of the R67 DHFR gene have been described (10). PCR reactions were carried out in the following reaction mixture: 10 mM Tris–HCl pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 100 pmol of each primer and 5 U *Taq* polymerase (Roche). The dNTP concentrations are described in the tables and legends. Input was ~5 ng plasmid DNA. The cycling parameters were: 50× (95°C, 30 s; 60°C 30 s; 72°C 10 min). Long elongation times were used to favour elongation after mismatches. Vent (Biolabs) and rTth (Roche) DNA polymerases were used at 2 and 2.5 U per reaction. $MnCl_2$ and dNTPs were purchased from Sigma and Pharmacia. PCR products were cloned via *Sac*I and *Bam*HI restriction sites and individual colonies picked, grown up and sequenced as described (10). A few products were cloned into the *Sac*I and *Bam*HI site of M13mp18 RF DNA. Recombinants were sequenced using thermosequenase (USB Amersham).

Unlike the *E.coli* chromosomal counterpart, the R67 DHFR gene is resistant to trimethoprim ($trim^R$). As the pTrc99A (Stratagene) cloning vector confers resistance to ampicillin ($ampi^R$) the ratio of the number of colonies on trimethoprim plus ampicillin and ampicillin only plates yields the proportion of functional genes post-PCR. The plating efficiencies of wild-type

* To whom correspondence should be addressed

DHFR construct on trimethoprim and ampicillin plates were comparable. Greater than 90% of ampi^R colonies had DHFR inserts.

RESULTS

DNA hypermutagenesis

Given the modified amplification protocol PCR conditions were first optimized for primer, magnesium, *Taq* DNA polymerase concentrations and number of cycles. As usual there was a strong Mg^{2+} dependence for all the thermostable polymerases used, the 2.5–5 mM range proving satisfactory. Particularly with large dNTP biases 30 cycles of PCR yielded relatively little product. Fifty cycles allowed adequate recovery for all but one reaction involving a 1000-fold [dTTP]/[dCTP] bias. In this case a further 25 cycles with equimolar dNTPs were performed as a chase. The efficiency of a standard amplification with equimolar 50 μM dNTPs was not affected by the addition of 1 mM ATP indicating that any increase in the ionic strength resulting from the addition of millimolar triphosphate did not alter PCR yields (data not shown).

Table 1 gives viable mutant frequencies following DNA hypermutagenesis with increasing [dTTP] > [dCTP] biases. The inverse relationship between the proportion of trim^R colonies as a function of the total (i.e. ampi^R) with increasing bias reflects the extent of DNA hypermutation. The overall mutation frequency for the entire amplification was inversely proportional to the dNTP bias and attained values as high as 2.9×10^{-2} substitutions per base per reaction for the ampi^R clones (Table 2). A collection of hypermutated trim^R sequences is given in Figure 1A. Up to five amino acid substitutions per functional clone (6.5%) were obtained which were generally well distributed throughout the sequence. Among the most hypermutated ampi^R clones up to 15 (6.5%) nucleotides and 11 (14%) amino acids respectively were replaced (not shown). The vast majority of substitutions were GC→AT transitions, as predicted from G_t:T mispairing on both strands due to the [dTTP] > [dCTP] bias. A small proportion (6%) of transversions were noted, uniquely A→T and T→A, to be expected from what is known about the ability of *Taq* DNA polymerase to elongate after mismatches (18,19).

Use of the thermostable rTth DNA polymerase, with reverse transcriptase activity, upon PCR with biased dNTP pools did not differ significantly from *Taq* DNA polymerase as judged by the ratio of trim^R and ampi^R clones and was not pursued. As a control DNA hypermutagenesis was performed using the Vent DNA polymerase which has a 3′-exonuclease activity. As evidenced by sequencing 20 trim^R clones, Vent completely protected DNA amplification against base misincorporation with a 500-fold [dTTP]/[dCTP] bias (data not shown).

Table 1. Inverse relationship between functional R67 DHFR mutants and dNTP pool biases

dNTP/μM C	T	A	G	R67 DHFR trim^R/ampi^R	
1000	1000	50	50	1100/1160	95%
10	1000	50	50	189/463	41%
3	1000	50	50	221/1373	16%
1	1000	50	50	42/420	10%

The [dTTP] > [dCTP] bias would favour GC→AT transitions. Plating of cloned PCR products on ampicillin (ampi^R) yields the total number of recombinants, while plating on trimethoprim and ampicillin (trim^R) yields the number of functional recombinants.

Table 2. DNA hypermutation frequencies

dNTP/μM C	T	A	G	Mn^{2+} mM	Colonies sequenced	No. mut.[a]	Ti/Tv[b]	N→A,T[c]	N→G,C[d]	Mutation frequency[e]
1	1000	50	50	–	37 trim^R	126	125/1	121	5	1.5×10^{-2}
1	1000	50	50	–	24 ampi^R	162	157/5	157	5	2.9×10^{-2}
3	1000	50	50	–	20 trim^R	24	22/2	24	0	5.2×10^{-3}
3	1000	50	50	–	20 ampi^R	29	27/2	28	1	6.3×10^{-3}
10	1000	50	50	–	20 trim^R	15	12/3	12	3	3.2×10^{-3}
10	1000	50	50	–	22 ampi^R	21	16/5	19	2	4.1×10^{-3}
5	1000	5	1000	–	18 trim^R	3	3/0	3	0	7.2×10^{-4}
5	1000	5	1000	–	18 ampi^R	0	0/0	0	0	$\leq 10^{-4}$
30	1000	30	1000	–	18 ampi^R	4	4/0	1	3	10^{-3}
30	1000	30	1000	0.5	34 ampi^R	755	521/234	256	499	10^{-1}
100	1000	100	1000	0.5	18 trim^R	19	16/3	8	11	4.5×10^{-3}
100	1000	100	1000	0.5	24ampiR	41	33/8	11	30	7.4×10^{-3}

On the left are the experimental conditions in terms of dNTP and manganese cation concentrations. On the right the analysis of mutants.

[a]Total number of mutations scored.

[b]Ti/Tv, number of transitions/number of transversions.

[c]Number of substitutions from non-A→A and non-T→T combined.

[d]Number of substitutions from non-G→G and non-C→C combined.

[e]The average mutation frequency was calculated as the number of substitutions divided by product of the number of clones sequenced and the number of bases between the PCR primers (231 bp).

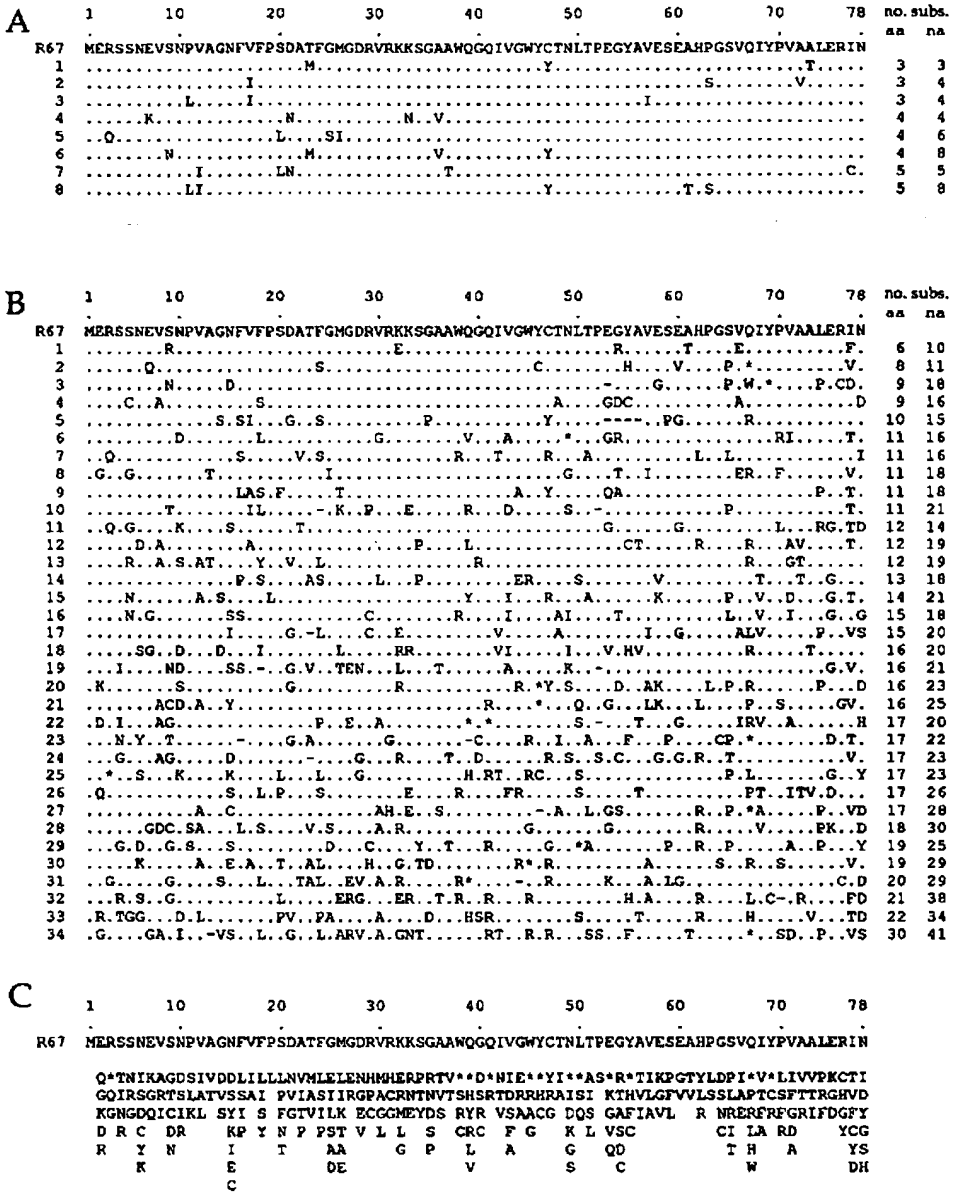

Figure 1. Collections of R67 DHFR hypermutants. Amino acid sequences were aligned to the reference R67 sequence using the one letter code, only differences being noted (10). A dot indicates sequence identity, a hyphen (-) represents codons harbouring nucleotide deletions. An asterisk (*) defines an in phase stop codon. To the left is the clone designation, to the right the number of amino acid (aa) and nucleic acid (na) differences with respect to the wild-type DHFR sequence. (**A**) A selection of trim^{R} hypermutants derived from the single bias [dTTP] = 1 mM, [dCTP] = 1 μM reaction (Tables 1 and 2). (**B**) Thirty-four clones derived from hypermutation involving two dNTP biases ([dTTP] = [dGTP] = 1 mM, [dCTP] = [dATP] = 30 μM) with 0.5 mM manganese cations (Table 2). All clones were trim^{S}. Among all data sets no two sequences were identical. Clone 33 and 34 encoded single nucleotide insertions at codons 13 and 31 (not shown). (**C**) Summary of all amino acid substitutions from all data sets. On average 3.7 substitutions per residue (range 1–7) were identified from a trivial number of clones. The single letter amino acid code is: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Double dNTP biases and manganese ions

The particularities of the G_t:T mismatch ensured A+T enrichment of the R67 DHFR gene. Alternatively a [dGTP] > [dATP] bias would have generated T_t:G mismatches with resulting G+C enrichment. Yet if all four base transitions could be generated during a single reaction the resulting mutant libraries would be among the most complex possible accessing an even greater proportion of sequence space. This is in principle possible if both a [dTTP] > [dCTP] and [dGTP] > [dATP] bias were used during PCR. However, no product whatsoever was obtained with a 1000- or 300-fold biases in both ratios. Only with <200-fold biases was

	A	T	C	G
A	-	0.21	0.05	0.90
T	0.56	-	1	0.06
C	0.02	0.28	-	0.06
G	0.31	0.04	0.08	-

Figure 2. Base substitution matrix for the 755 mutation data set normalized to the most frequent substitution, T→C. Substitution frequencies (e.g. G→A/755) given with respect to the (+) DNA strand were normalized to its base composition (52 T, 60 C, 65 G and 54 A excluding the ATG initiator codon derived from the forward PCR primer).

this possible. Sequencing of the $trim^R$ hypermutated products yielded unexpectedly low mutation frequencies (Table 2).

Transition metal ions such as manganese (Mn^{2+}) and cobalt (Co^{2+}) may decrease the fidelity of DNA synthesis including PCR (12,20,21). Addition of $MnCl_2$ to a final concentration of 0.5 mM in a reaction with both [dTTP]/[dCTP] = [dGTP]/[dATP] = 1000 μM/30 μM overcame the enhanced fidelity noted above. The overall base mutation frequency could be increased from ~10^{-3} to ~10^{-1} per site per amplification (Table 2). In fact, the PCR was so error prone that no $trim^R$ colonies (0 $trim^R$/600 $ampi^R$) were identified. A collection of 34 clones is given in Figure 1B, mutants starting with a minimum of 10 substitutions (4%) per clone. The maximum number was 41 (18%) per clone. The proportion of transversions (31%) was greatly enhanced by the addition of Mn^{2+} and was accompanied by a few deletions and even fewer single base insertions (Fig. 1B). There was no correlation between the proportion of synonymous (s) to non-synonymous (ns) base substitutions within this or any other data sets (not shown).

Figure 1C collates amino acid replacements from all the data sets and indicates that hypermutagenic PCR may introduce between one and seven (mean 3.7) different amino acids per residue. The large 755 mutation data set resulting from manganese mutagenesis was analyzed for substitution biases. The mutation matrix, normalized for base composition effects, showed almost perfect strand symmetry (i.e. G→C ≈ C→G, etc.) (Fig. 2). However, there was a bias for AT→GC transitions which perhaps may be attributable to subtle differences between G_t:T and T_t:G mismatches in the *Taq* DNA polymerization site. Once again, A→T and T→A were the most frequent transversions.

The proportions of adjacent double and triple substitutions were as expected given the mutation frequency of isolated changes. The distribution of substitutions per site differed significantly from that expected from a binomial distribution (not shown). A χ^2 analysis of the dinucleotide context showed that there were a few substitution preferences, notably an excess of AT→GC transitions as well as A→T and T→A transversions in the context of CpA/TpG, and a dearth of the same transitions in the TpA dinucleotide. Only a one significant bias (GC→AT in GpC) was seen in reactions with [dTTP] > [dCTP] biases.

DISCUSSION

Balanced DNA precursor concentrations are clearly crucial to the fidelity of cellular DNA or retroviral cDNA synthesis *in vivo* and *in vitro* (22–25). The same is true of PCR, the present findings reproducing and extending earlier work (12,13,16). The nature of the dNTP bias generally produced the substitution expected from G:T mispairing once again highlighting the importance of this most stable of base mismatches to hypermutation (7,8). Perhaps surprisingly, the fidelity of amplification was enhanced many fold when both deoxypyrimidine and deoxypurine triphosphate biases were used (Table 2). This might result from the fact that although G:T mismatches are being forced so were G:G and T:T mispairs. From what is known of *Taq* DNA polymerase elongation beyond mismatches, G:G represents one of the most substantial blocks to elongation and consequently amplification (18,19). By contrast T:T mismatches pose fewer problems. The addition of Mn^{2+} ions, known to be mutagenic for DNA synthesis by a variety of mechanisms including modification of the relative K_ms of mismatches and matches (20,21), overcame this problem. The 100-fold enhanced overall mutation frequency was indeed so great that no $trim^R$ clones could be derived.

With a double dNTP bias and manganese ions there was an excess of transitions towards G+C which was not strand-specific (Fig. 2). Clearly this could be countered by increasing [dTTP] or decreasing [dGTP] in the reaction. There was evidence that the distribution of mutations was not completely random. However, significant deviations from the expected values were noted for only a few substitutions.

A comparison of RNA and DNA hypermutagenesis is telling (7,8). The HIV-1 reverse transcriptase error rate per pass is clearly greater than *Taq* DNA polymerase. Among the hundreds of RNA molecules hypermutated *in vitro* by the HIV-1 reverse transcriptase, up to 32% of G targets were substituted for one clone with a best mean of 11%, all in a single cycle of cDNA synthesis (7). However, given the monotony (e.g. G→A) of RNA hypermutagenesis these numbers translates into best and average overall mutation frequencies of ~7 and 3% respectively. To date, DNA hypermutagenesis has produced up to 18% base substitution per clone with a best mean of 10% involving copying of both strands.

Despite the intrinsic properties of the HIV-1 RT the advantages of DNA hypermutagenesis by PCR are manifold. First, the complexity of the mutant libraries are incomparably greater providing access to even larger fraction of sequence space. Secondly, the procedure is faster being reduced to a single reaction. Thirdly, as the PCR step is mutagenic there is in principal no need to clone before undertaking a second cycle of DNA hypermutagenesis. However, the power of DNA hypermutagenesis is now so great that iteration without some sort of phenotypic selection is probably unwise because the information threshold can be crossed. In addition, preliminary work suggests that primer dimers and deleted molecules may be preferentially amplified upon cycling without phenotypic selection or purification of the DNA band. The conditions can surely be refined to purge the present GC→AT bias.

The extent of mutation described above, as well as the complexity of the mutant libraries, exceeds that generated by any biological method to date. A recent paper described hypermutagenic PCR using modified dCTP and dGTP substrates (26). The best and average mutation frequencies described here

(0.18 and 0.1 per base per reaction) are highly comparable with those reported, notably 0.19 and 0.1 per base per reaction. The modified bases generally produced AT→GC transitions and a small percentage (<10%) of transversions. The present protocol used standard bases, generates at high frequencies all four transitions and, given the presence of manganese cations, approximately one third transversions. Clearly there is considerable flexibility and choice in the production of hypermutants which could be tailored to the desires or needs of the experimentalist.

Although DNA hypermutagenesis allows huge leaps through sequence space, viable hypermutants are to be had. The diversity currently accessible is so great that any screening procedure will explore only a minute fraction of the sequence space accessed. The simplicity and efficiency of DNA hypermutagenesis transfers the burden of work in protein evolution *in vitro* onto analytical procedures. The potential of the method is such that, after iterative DNA hypermutagenesis, the historical information content of a sequence might be annihilated, defying recognition.

The choice of the small DHFR gene was particularly propitious. PCR product yield decreases with dNTP pool bias and is further reduced upon addition of Mn^{2+} cations. This can be alleviated to some extent by a chase PCR with equimolar dNTPs. Alternatively cycling the product from an agarose gel purified band should allow one to extensively hypermutate larger genes. Yet as the probability of introducing deleterious mutations increases with target DNA length, inevitably hypermutation of such genes might not prove as informative, unless some form of biological selection is used. A further reservation concerns the nature of the transversions observed. That A→T and T→A transitions were the most common may be attributed to the ability of *Taq* DNA polymerase to elongate after T:T mismatches (18,19). Inversely, the dearth of a number transitions correlates well with the relative inefficiency of the enzyme to elongate after A_t:G, G_t:A, G_t:G and C_t:C mismatches. Thus the mutation spectrum is shaped to some extent by *Taq* DNA polymerase. It is possible that different thermostable enzymes might show subtle differences. Alternatively, modifications to the reaction mix might be introduced in an attempt to alleviate such preferences.

DNA hypermutation accelerates what may occur under more physiological circumstances over much longer time periods. Indeed there is a wealth of experimental data associating dNTP pool biases, mutation and cancer (22,23,25). The consequences of an intracellular [dTTP] > [dCTP] bias are particularly intriguing. Among eukaryotic cells the intracellular dNTP concentrations are invariably [dATP] ≥ [dTTP] > [dCTP] ≥ [dGTP] or, in other words, [dTTP] > [dCTP] and [dATP] > [dGTP] (25). Given the particular properties of the G:T mismatch any increase in the deoxypyrimidine triphosphate bias would help enrich the sequence in A+T. The potential mutagenic effects resulting from fluctuations in the deoxypurine triphosphate bias would have to be even more substantial as they would need to invert the natural [dATP] > [dGTP] bias (25). From this it might be surmised that any exacerbation of the natural [dTTP] > [dCTP] bias should have more long term impact on the genome. In this context it is interesting to note that among vertebrate cells non-coding segments are generally A+T rich.

It is salutary to realize that DNA synthesis can be so error prone. It might be supposed that during the evolution of primitive DNA based replicons and before highly integrated dNTP metabolism, biased dNTP concentrations alone, or in conjunction with dilute solutions of some transition metal ions, might have contributed to the genesis of DNA sequence diversity upon which natural selection could work.

ACKNOWLEDGEMENTS

We would like to thank Drs Fredj Tekaia and Christophe Terzian for statistical analyses. This work was supported by grants from Institut Pasteur and l'Agence Nationale pour la Recherche sur le SIDA.

REFERENCES

1 Drake, J. W. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 4171–4175.
2 Holland, J. J., Domingo, E., de la Torre, J. C. and Steinhauer, D. A. (1990) *J. Virol.*, **64**, 3960–3962.
3 Rueda, P., Garcia-Barreno, B. and Melero, J. A. (1994) *Virology*, **198**, 653–662.
4 Pathak, V. K. and Temin, H. M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 6019–6023.
5 Vartanian, J. P., Meyerhans, A., Åsjö, B. and Wain-Hobson, S. (1991) *J. Virol.*, **65**, 1779–1788.
6 Vartanian, J. P., Meyerhans, A., Sala, M. and Wain-Hobson, S. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3092–3096.
7 Martinez, M. A., Vartanian, J. P. and Wain-Hobson, S. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 11787–11791.
8 Martinez, M. A., Sala, M., Vartanian, J. P. and Wain-Hobson, S. (1995) *Nucleic Acids Res.*, **14**, 2573–2578.
9 Pezo, V., Martinez, M. A. and Wain-Hobson, S. (1996) *Nucleic Acids Res.*, **24**, 253–256.
10 Martinez, M. A., Pezo, V., Marliere, P. and Wain-Hobson, S. (1996) *EMBO J.*, **15**, 1203–1210.
11 Battula, N. and Loeb, L. A. (1974) *J. Biol. Chem.*, **249**, 4086–4093.
12 Leung, D., Chen, E. and Goeddel, D. (1989) *Technique*, **1**, 11–15.
13 Caldwell, R. C. and Joyce, G. F. (1992) *PCR Meth. Appl.*, **2**, 28–33.
14 Chen, K. and Arnold, F. H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5618–5622.
15 Rellos, P. and Scopes, R. K. (1994) *Prot. Eng. Purif.*, **5**, 270–277.
16 Fromant, M., Blanquet, S. and Plateau, P. (1995) *Anal. Biochem.*, **224**, 347–353.
17 Modrich, P. (1987) *Annu. Rev. Biochem.*, **56**, 435–466.
18 Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. (1990) *Nucleic Acids Res.*, **18**, 999–1005.
19 Huang, M. M., Arnhein, N. and Goodman, M. F. (1992) *Nucleic Acids Res.*, **20**, 4567–4573.
20 Beckman, R. A., Mildvan, A. S. and Loeb, L. A. (1985) *Biochemistry*, **24**, 5810–5817.
21 Goodman, M. F., Keener, S., Guidotti, S. and Branscomb, E. W. (1983) *J. Biol. Chem.*, **258**, 3469–3475.
22 Phear, G., Nalbantoglu, J. and Meuth, M. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 4450–4454.
23 Meuth, M. (1989) *Expt. Cell. Res.*, **181**, 305–316.
24 Kunz, B. and Kohalmi, S. (1991) *Annu. Rev. Genet.*, **25**, 339–359.
25 Mathews, C. K. and Ji, J. (1992) *BioEssays*, **14**, 295–301.
26 Zaccolo, M., Williams, D. M., Brown, D. M. and Gherardi, E. (1996) *J. Mol. Biol.*, **255**, 589–603.